

Il problema della negazione nel cervello

Assenza, vuoto intenzionale e il paradosso della rappresentazione del non-essere

Il paradosso e le sue radici logiche

I. Un sistema che non può tacere il falso

C'è un momento, nella storia di ogni teoria della mente, in cui un problema apparentemente tecnico si rivela essere il nucleo di tutto. Per le neuroscienze cognitive, questo momento è forse ancora davanti a noi — e il problema che lo annuncia è quello della negazione. Non la negazione come operatore sintattico, non il morfema *non* come categoria lessicale, ma la negazione come operazione cognitiva fondamentale: la capacità di un sistema biologico di rappresentare ciò che non è, di tenere presente l'assente, di produrre e manipolare la forma logica della non-verità.

Il paradosso è il seguente. Nella sua struttura biofisica più elementare, il cervello è un sistema di attivazione: i neuroni si eccitano, le sinapsi si potenziano, le popolazioni convergono verso stati di scarica collettiva. Tutto ciò che il cervello fa, lo fa attraverso l'attivazione — attraverso il *fare accadere qualcosa* a livello di potenziali di membrana, neurotrasmettitori, campi elettrici. Anche l'inibizione, che potrebbe sembrare l'eccezione, è un'attivazione: gli interneuroni inibitori *scaricano*, i recettori GABA *si legano*, i canali del cloro *si aprono*. L'inibizione è un processo fisico positivo, non l'assenza di processo.

Eppure il cervello umano pensa la negazione. Afferma che Cesare non è vivo, che il triangolo non è rotondo, che domani non pioverà. Costruisce dimostrazioni per assurdo, in cui assume P per derivarne $\neg P$ e concludere $\neg P$. Immagina controfattuali — *se avessi preso l'altra strada* — che rappresentano esplicitamente stati del mondo che non si sono verificati. Comprende la differenza tra *Mario non ha mangiato* e *non è vero che Mario ha mangiato*, e tra entrambe e *Mario ha smesso di mangiare*, che presuppone che Mario abbia mangiato in passato mentre nega che lo faccia ora. Tutte queste operazioni richiedono la rappresentazione proposizionale dell'assenza — non la semplice mancanza di attivazione, ma l'attivazione di una struttura che *indica* la mancanza come contenuto.

Come si realizza questo, in un sistema che può soltanto attivare?

Questa domanda non ha ancora ricevuto una risposta coerente e formalmente articolata nella letteratura neuroscientifica. Vi sono studi sull'elaborazione delle frasi negative, dati ERP sulle difficoltà di comprensione del *non*, proposte frammentarie sulla codifica dell'errore di previsione come forma di negazione implicita. Ma manca una teoria — una teoria nel senso pieno del termine, con un nucleo formale, implicazioni sperimentali testabili, e una collocazione precisa nel dibattito filosofico sulla natura della negazione. Questo saggio tenta di costruirne il prodromo.

II. La negazione nella logica classica e il suo imbarazzo computazionale

La negazione della logica classica — il connettivo $\neg$ della logica proposizionale e del calcolo dei predicati — è definita dalla sua tavola di verità: $\neg P$ è vero quando P è falso, e falso quando P è vero. Questa definizione è elegante, bivalente, e computazionalmente trattabile: un circuito logico che implementa la porta NOT produce il complemento di qualunque bit in ingresso in un singolo ciclo di clock.

Ma questa eleganza nasconde un'assunzione profonda e problematica: che la negazione operi su *valori di verità* determinati, non su *proposizioni nel corso della loro valutazione*. La logica classica è una logica del già-valutato: ogni proposizione ha già, per ogni interpretazione, un valore di verità definito, e la negazione si limita a invertirlo. Non c'è processo, non c'è tempo, non c'è incertezza: solo la struttura atemporale della funzione booleana.

Il cervello, per contro, non opera su valori di verità già determinati: opera su rappresentazioni in costruzione, su inferenze in corso, su previsioni che si aggiornano in tempo reale. La proposizione "*Mario è entrato nella stanza*" non ha, per il cervello dell'ascoltatore, un valore di verità preassegnato: viene valutata dinamicamente sulla base delle evidenze disponibili — il rumore di passi udito, la porta che si è aperta, il modello del discorso corrente, le aspettative sul comportamento di Mario. Quando il sistema cognitivo produce la negazione di questa proposizione, non sta invertendo un bit preassegnato: sta facendo qualcosa di computazionalmente molto più complesso e teoricamente molto meno compreso.

La prima difficoltà è quella della **scope** — la portata della negazione. In logica classica, $\neg P$ ha portata sull'intera proposizione P . Ma nel linguaggio naturale, la negazione può avere portata selettiva su diversi elementi della proposizione: "*Non ho mangiato la torta*" può negare l'intera proposizione (non è accaduto che io mangiassi la torta), o può negare soltanto il complemento oggetto (ho mangiato qualcosa, ma non la torta — negazione focalizzata), o può negare il soggetto (qualcuno ha mangiato la torta, ma non io). Queste varianti semantiche non sono equivalenti e vengono disambiguate, nel parlato naturale, attraverso l'accento prosodico e la struttura informativa. Il cervello deve dunque determinare la

portata della negazione prima di poterne computare il significato — e la determinazione della portata richiede l'integrazione di informazioni prosodiche, pragmatiche e contestuali che vanno ben al di là della struttura sintattica locale.

La seconda difficoltà è quella della **presupposizione**. Le frasi negative non sono semanticamente simmetriche rispetto alle loro controparti affermative: "*Mario ha smesso di fumare*" presuppone che Mario fumasse in precedenza, e la sua negazione "*Mario non ha smesso di fumare*" mantiene questa presupposizione — nega soltanto l'interruzione, non il fumare. La struttura presuppositiva non è invertita dalla negazione: è, come direbbe Frege, parte del *sense* dell'espressione che rimane stabile attraverso la negazione, non parte della sua *denotazione* che viene invertita. Questa distinzione — tra contenuto proposizionale negabile e presupposizione non-negabile — è stabile nel linguaggio naturale ma non ha una rappresentazione ovvia nei sistemi neurali che elaborano le frasi.

La terza difficoltà, forse la più profonda, è quella della **negazione metalinguistica**. In alcuni contesti, la negazione non nega il contenuto proposizionale ma la scelta lessicale o il livello di descrizione: "*Non è un cane, è uno schnauzer*" non afferma che l'animale in questione non sia un cane nel senso biologico — è certamente un cane — ma protesta contro la scelta di denominarlo *cane* anziché *schnauzer*, implicando che il livello di specificità della descrizione sia inappropriato. Questa negazione metalinguistica, analizzata da Laurence Horn, non è un'operazione sulla proposizione ma un'operazione sulla forma della proposizione stessa — un atto che si svolge al livello del codice, non al livello del messaggio. Come il cervello distingue la negazione proposizionale dalla negazione metalinguistica è un problema che non ha ancora ricevuto risposta neuroscientifica.

III. La negazione intuizionista: dal valore di verità al processo di prova

La logica intuizionista — sviluppata da Luitzen Egbertus Jan Brouwer come fondamento di una matematica costruttivista, e formalizzata da Arend Heyting nel suo calcolo proposizionale del 1930 — offre una concezione della negazione radicalmente diversa da quella classica, e molto più congeniale all'implementazione in sistemi computazionali temporalmente estesi come il cervello.

Nella logica intuizionista, il significato di una proposizione non è il suo valore di verità ma la descrizione dei *metodi di prova* che la verificherebbero. Una proposizione è vera se e soltanto se si è in possesso di una prova di essa — non se è vera in qualche mondo platonico di valori di verità preassegnati, ma se si può costruire un oggetto matematico che costituisce la sua dimostrazione. Questa concezione della verità è **costruttiva** e **procedurale**: non riguarda stati, ma processi.

In questo framework, la negazione $\neg P$ è definita non come l'affermazione che P è falso, ma come la derivazione di un'assurdità da P : $\neg P$ significa $P \rightarrow \perp$, dove $\perp$ è la proposizione assurda (l'esplosione, l'inconsistenza). Negare P significa mostrare che qualunque presunta prova di P conduce alla contraddizione — ossia che il *processo* di costruzione di una prova di P , se portato a termine, produce un oggetto incoerente.

Questa definizione ha una conseguenza notevole: in logica intuizionista, il **terzo escluso** — la tautologia classica $P \vee \neg P$ — non è valido in generale. Non ogni proposizione è tale che o se ne possiede una prova o se ne possiede una confutazione: per molte proposizioni matematiche, la situazione è quella dell'ignoranza genuina — né P né $\neg P$ sono dimostrabili con i metodi attualmente disponibili. La logica intuizionista è dunque epistemicamente onesta rispetto all'incompletezza della conoscenza, laddove la logica classica assume tacitamente l'onniscienza.

Per il cervello, questa concezione della negazione è computazionalmente molto più naturale della versione classica. Il cervello non ha accesso ai valori di verità delle proposizioni sul mondo: ha accesso alle *evidenze* che supportano o indeboliscono diverse ipotesi sul mondo, e aggiorna le proprie rappresentazioni in funzione di queste evidenze. In questo senso, il cervello è intrinsecamente intuizionista: opera in un regime di incertezza epistemica in cui né P né $\neg P$ sono mai pienamente determinate, e la negazione è sempre una negazione di processo — la segnalazione che il processo di costruzione della rappresentazione P ha condotto a una contraddizione con le evidenze disponibili.

Il formalismo intuizionista che cattura questa struttura con maggiore precisione è la **semantica di Kripke** per la logica intuizionista. Un **frame di Kripke** è una coppia $(W, \leq)$ dove W è un insieme di *stati epistemicamente possibili* — le diverse configurazioni di informazione disponibile all'agente — e $\leq$ è una relazione di accessibilità che riflette l'ordine di accumulazione dell'informazione: $w \leq v$ significa che lo stato v è epistemicamente più ricco di w (contiene almeno tante informazioni). La funzione di valutazione V assegna a ogni coppia (P, w) un valore in $\{0, 1\}$ con la condizione di **persistenza**: se $V(P, w) = 1$ e $w \leq v$, allora $V(P, v) = 1$ — una volta che una proposizione è stata dimostrata, rimane dimostrata in tutti gli stati epistemicamente successivi.

La negazione intuizionista nella semantica di Kripke è:

$$V(\neg P, w) = 1 \text{ sse per ogni } v \geq w, V(P, v) = 0$$

Questa condizione è molto più esigente della negazione classica: $\neg P$ è vera nello stato w non soltanto se P è attualmente non-dimostrata, ma se P è non-dimostrabile in nessun futuro stato epistemicamente accessibile da w — se nessuna accumulazione futura di informazione potrebbe mai produrre una prova

di P. La negazione intuizionista è dunque una negazione *robusta*, che afferma non solo l'assenza corrente di prova ma l'impossibilità strutturale di prova.

Questa struttura corrisponde sorprendentemente bene alla fenomenologia della negazione nella cognizione naturale. Quando si dice "*questo non è un triangolo*", non si afferma semplicemente che attualmente non si possiede una prova della triangolarità dell'oggetto — si afferma che nessuna osservazione futura potrebbe produrre tale prova, perché l'oggetto ha proprietà incompatibili con la definizione di triangolo. Questa è negazione robusta nel senso della semantica di Kripke. Per contrasto, quando si dice "*non so se pioverà domani*", si nega non la proposizione ma la propria competenza epistemica corrente — una negazione non-robusta, compatibile con l'acquisizione futura di informazioni meteorologiche che la risolvono.

Il cervello distingue questi due tipi di negazione — e questa distinzione ha correlati neurali precisi che cominciano ad emergere dalla letteratura ERP, come si argomenterà nella sezione successiva. Ma la struttura formale che li distingue — la semantica di Kripke, con la sua differenza tra negazione debole (mancanza attuale di prova) e negazione forte (impossibilità strutturale di prova) — non era mai stata proposta come cornice teorica per la neuroscienza della negazione. Questo è uno dei contributi che questo saggio intende offrire.

IV. La negazione lineare: risorse, consumo e la logica del metabolismo neurale

Una terza concezione della negazione, meno nota della classica e dell'intuizionista ma teoricamente feconda per le neuroscienze, è quella della **logica lineare**, sviluppata da Jean-Yves Girard nel 1987. La logica lineare è una logica delle **risorse**: le proposizioni non sono verità atemporali ma *token di risorse* che possono essere usate esattamente una volta. L'operazione logica di inferenza *consuma* le proposizioni premessa, che non sono più disponibili per usi successivi.

In questo framework, la negazione $A \perp$ di una proposizione A non è la sua falsificazione ma la sua *duale*: l'oggetto che, combinato con A , produce il token di trivialità $\perp$. La metafora guida è quella della coniugazione lineare: A e $A \perp$ sono come un debito e il suo creditore — la loro combinazione si annulla. La negazione non afferma che qualcosa non è: afferma che qualcosa è il *complemento duale* di qualcos'altro.

Questa concezione ha un'analogia biologica sorprendentemente precisa nella biochimica del sistema nervoso. I neurotrasmettitori eccitatori e inibitori — glutammato e GABA — funzionano come coppia duale: il glutammato depolarizza, il GABA iperpolarizza; la loro combinazione può produrre l'equilibrio del potenziale di membrana. Il ciclo eccitazione-inibizione che genera le oscillazioni corticali discusse nel saggio precedente è, in un senso letterale, una dinamica di coppia duale nel senso della logica lineare. Il cervello

non rappresenta la negazione *aggiungendo* un segnale negatorio al segnale affermativo: la realizza attraverso la *dualità* tra i due tipi di input — una dualità che è, biofisicamente, il cuore dell'architettura corticale.

La connessione tra logica lineare e biologia neurale non è soltanto metaforica. Lafont e collaboratori hanno proposto, in un'analisi rimasta marginale rispetto all'attenzione che meritava, che la struttura della logica lineare sia isomorfa alla struttura delle reti di Petri — formalismi per la modellazione di sistemi concorrenti con risorse finite — e che le reti di Petri siano a loro volta un modello naturale per i circuiti biochimici cellulari, inclusi i circuiti di regolazione genica e i circuiti sinaptici. Se questa catena di isomorfismi regge, la logica lineare non è una teoria astratta applicata per analogia alla biologia: è la struttura formale intrinseca dei processi biologici computazionali.

Per la negazione, l'implicazione più interessante della logica lineare riguarda la distinzione tra **negazione reversibile** (why not, operatore $\neg$) e **negazione irreversibile** (bang, operatore $!$). La prima permette di *riusare* una risorsa negata; la seconda la consuma definitivamente. Nel sistema nervoso, questa distinzione corrisponde alla differenza tra inibizione reversibile — l'inibizione sinaptica che può essere superata da input eccitatori sufficienti — e inibizione irreversibile — l'inibizione che produce modificazioni permanenti della connettività, come la depressione a lungo termine (LTD) che riduce permanentemente l'efficacia sinaptica. La negazione che il cervello produce quando apprende che una credenza è falsa non è la stessa operazione della negazione momentanea che si produce durante l'elaborazione di una frase negativa: la prima modifica il paesaggio energetico della rete, consumando irreversibilmente la risorsa della credenza precedente; la seconda è una modulazione reversibile che non lascia traccia permanente.

Questa distinzione — tra negazione reversibile e negazione irreversibile, tra inibizione funzionale e plasticità a lungo termine — è una delle più importanti della psicologia cognitiva, e la logica lineare offre gli strumenti formali per trattarla con la precisione che merita.

V. Il silenzio di Frege e la domanda rimossa

C'è un momento nell'opera di Frege in cui il problema della negazione affiora nella sua radicalità e viene — in modo significativo — aggirato. Nel saggio *Die Verneinung* (1919), Frege si chiede se la negazione sia un'operazione sul *sensu* della proposizione o sul suo *valore di verità*, e conclude che la proposizione negata e la proposizione affermata hanno lo stesso senso — esprimono lo stesso *pensiero* — ma valori di verità opposti. La negazione non cambia ciò di cui si parla; cambia la valutazione di ciò di cui si parla.

Questa posizione è coerente con l'architettura semantica fregeana — la distinzione tra senso (il modo di presentazione) e denotazione (il riferimento o il

valore di verità) — ma lascia senza risposta una domanda che Frege non pone esplicitamente: in che relazione cognitiva si trova il soggetto pensante con una proposizione negata? Affermare P e affermare $\neg P$ implicano lo stesso pensiero, nella terminologia fregeana, ma implicano atteggiamenti proposizionali diversi — o meglio, implicano che il soggetto si trovi in stati cognitivi diversi rispetto allo stesso contenuto proposizionale.

Questo problema — il problema dell'*atteggiamento* verso la proposizione negata, distinto dal contenuto della proposizione stessa — è il punto in cui la semantica fregeana incontra i suoi limiti e in cui la fenomenologia diventa indispensabile. Il soggetto che pensa " P " e il soggetto che pensa " $\neg P$ " sono in relazione con lo stesso *pensiero* fregeano, ma si trovano in stati mentali profondamente diversi: il primo è in uno stato di asserzione, con le conseguenti aspettative e disposizioni all'azione; il secondo è in uno stato di negazione, con aspettative e disposizioni diverse e parzialmente opposte. Come si rappresenta neurologicamente questa differenza di atteggiamento proposizionale, distinta dalla differenza di contenuto?

Husserl, che a differenza di Frege non separava il contenuto proposizionale dall'atto intenzionale che lo presenta, aveva posto questa domanda in modo esplicito nelle *Ricerche Logiche* (1900–1901), distinguendo tra l'*atto* di giudizio e il suo *contenuto*, e mostrando che la negazione trasforma l'atto — da asserzione a rifiuto, da conferma a rigetto — senza necessariamente trasformare il contenuto nella stessa misura. La negazione è, per Husserl, primariamente una modificazione dell'atteggiamento intenzionale, non del contenuto rappresentato — e questa modificazione ha caratteristiche fenomenologiche proprie che la distinguono da qualunque altro atto cognitivo.

Nella parte seguente, si svilupperà questa intuizione husserliana in direzione di una teoria neurologicamente fondata del vuoto intenzionale — la struttura cognitiva attraverso cui il cervello non soltanto elabora le proposizioni negative ma le *vive* come tali, con il loro peculiare carattere di presenza dell'assente.

Il vuoto intenzionale, l'errore di previsione e il substrato neurale

VI. La presenza dell'assente: Husserl e il vuoto intenzionale

La fenomenologia husserliana affronta il problema della negazione da una prospettiva che non ha equivalente nella tradizione logico-analitica: la prospettiva dell'*atto intenzionale* nella sua struttura temporale vissuta. Per Husserl, ogni atto cosciente è intenzionale — si dirige verso un oggetto — e questa direzionalità non è una relazione tra un soggetto e un oggetto già dato, ma una struttura di *presentazione*: l'atto costituisce il proprio oggetto come tale, lo dà nella modalità specifica che gli è propria.

La negazione, in questa prospettiva, non è soltanto un operatore che inverte il valore di verità di una proposizione: è una **modalità di presentazione** specificamente diversa dall'affermazione. Quando si percepisce che il bicchiere atteso non è sul tavolo — che il bicchiere *manca* — non si assiste semplicemente all'assenza di un'attivazione visiva. Si vive una struttura intenzionale positiva: la presenza dell'assente come assente, il vuoto che occupa il luogo dove qualcosa avrebbe dovuto trovarsi e non si trova. Husserl chiama questa struttura *intenzione vuota* (*leere Intention*): un atto intenzionale che si dirige verso il suo oggetto senza che l'oggetto si dia nella presenza sensibile, ma che è nondimeno un atto positivo — non la semplice mancanza di percezione, ma la percezione della mancanza.

La differenza tra l'intenzione vuota e la semplice assenza di percezione è fenomenologicamente nettissima. Se sono in una stanza buia e non vedo il tavolo, non ho un'intenzione vuota verso il tavolo: ho semplicemente un campo percettivo privo di stimoli visivi relativi al tavolo. Ma se entro in una stanza aspettandomi di trovare il tavolo sul quale avevo lasciato i miei occhiali, e il tavolo non c'è, *vivo* l'assenza del tavolo come assenza — come la presenza di un vuoto con la forma del tavolo, localizzato nel punto dello spazio dove il tavolo avrebbe dovuto essere. Questa intenzione vuota è un atto cognitivo positivo, con contenuto intenzionale determinato e con conseguenze pratiche precise: cerco il tavolo, chiedo dove sia andato a finire, riorganizzo le mie aspettative spaziali.

Nelle *Idee per una fenomenologia pura* (1913), Husserl distingue tra la **posizione** dell'oggetto intenzionale — il fatto di esso come esistente, l'asserzione della sua realtà — e la sua **modificazione** attraverso la negazione. La negazione non abolisce l'intenzione verso l'oggetto: la modifica nella sua modalità positiva, trasformandola da posizione (asserzione dell'esistenza) a *posizione cancellata* (*durchgestrichene Setzung*): l'oggetto è ancora inteso — è ancora il correlato dell'atto intenzionale — ma è inteso nella modalità del *non-essere*, della presenza sbarrata.

Questa struttura ha una conseguenza che si rivelerà cruciale per l'analisi neurologica: il contenuto della proposizione negata deve essere *completamente rappresentato* prima che la negazione possa operare su di esso. Non si può negare ciò che non è stato prima affermato, almeno implicitamente. La negazione è logicamente, cognitivamente e — si argomenterà — neurologicamente *secondaria* rispetto all'affermazione: presuppone la costruzione della rappresentazione che poi nega.

Questa asimmetria temporale tra affermazione e negazione — che Husserl descrive in termini di stratificazione degli atti intenzionali — ha una controparte empirica precisa nella letteratura neurolinguistica. L'elaborazione di frasi negative è sistematicamente più lenta e più costosa in termini di risorse cognitive rispetto all'elaborazione delle corrispondenti frasi affermative, anche quando lunghezza e frequenza lessicale sono controllate. Questa asimmetria non è soltanto un fatto del linguaggio: riflette la struttura degli atti cognitivi sottostanti, in cui la negazione si costruisce sopra la rappresentazione affermativa che precede.

VII. L'errore di previsione come negazione biologica primaria

Il framework della codifica predittiva, introdotto nella discussione sul linguaggio, offre un punto di accesso privilegiato alla neurobiologia della negazione — e lo fa attraverso un concetto che è già parte del framework: **l'errore di previsione**. Nella gerarchia predittiva corticale, un errore di previsione emerge quando l'input sensoriale non corrisponde alla previsione generata dai livelli superiori. Questo errore — codificato come segnale gamma ascendente — è, nella sua struttura formale, una forma di negazione: è il segnale che $\neg P$, dove P è la previsione.

Ma l'identità tra errore di previsione e negazione non è immediata e richiede di essere argomentata con cautela. L'errore di previsione afferma che la previsione P non è stata confermata dall'input — che l'input è diverso da P . Ma questo non equivale alla proposizione $\neg P$ nel senso logico pieno: afferma soltanto che P non corrisponde all'input osservato, non che P sia falsa in senso assoluto. La differenza è quella tra una negazione *relativa* — negazione rispetto all'evidenza disponibile — e una negazione *assoluta* — negazione del contenuto proposizionale indipendentemente dall'evidenza.

Eppure questa distinzione, nella pratica cognitiva, è meno netta di quanto appaia. La negazione assoluta, quella che il soggetto afferma quando dice "*il triangolo non è rotondo*", è costruita sulla base di un'evidenza — la definizione geometrica del triangolo — che rende l'evidenza contraria strutturalmente impossibile. La negazione è assoluta perché l'errore di previsione sarebbe *massimo e garantito*: qualunque input presentato come "triangolo rotondo" produce un segnale di errore che non può essere ridotto da nessun aggiornamento del modello compatibile con la geometria euclidea. La negazione assoluta è l'errore di previsione *a priori* — l'errore che il sistema sa già di dover produrre senza aver bisogno di incontrare l'input.

Questa proposta — la negazione proposizionale come generalizzazione dell'errore di previsione dal dominio percettivo al dominio concettuale — suggerisce una gerarchia di tipi di negazione corrispondente a una gerarchia di livelli nella gerarchia predittiva:

Al livello più basso, la negazione percettiva primaria: l'errore di previsione sensoriale, il segnale gamma che comunica la discrepanza tra input e previsione. Questo è il mattone biologico della negazione — il tipo di negazione che condividiamo con i vertebrati superiori in generale.

Al livello intermedio, la negazione epistematica: la rappresentazione esplicita che una credenza o aspettativa si è rivelata errata, prodotta dall'integrazione degli errori di previsione attraverso i livelli superiori della gerarchia. Questo tipo di negazione richiede un modello esplicito delle proprie credenze — una forma

rudimentale di metacognizione — e probabilmente rappresenta la soglia filogenetica al di sopra della quale la negazione si fa proposizionale.

Al livello superiore, la negazione logica piena: l'operazione che nega una proposizione *indipendentemente dall'esperienza corrente*, attraverso la sola struttura del modello generativo. Questo tipo di negazione — il solo che la logica formale tratta — richiede la capacità di operare su rappresentazioni astratte sganciate dall'input immediato, e corrisponde al livello più alto della gerarchia predittiva, dove le rappresentazioni sono troppo astratte per avere un diretto correlato sensoriale.

Questa stratificazione può essere formalizzata nel seguente modo. Si consideri la gerarchia predittiva a L livelli, con stati latenti $x^{(l)}$ e errori di previsione $\varepsilon^{(l)}$ ad ogni livello. La negazione di tipo l è definita come:

$N^{(l)}[P] \equiv$ "P produce $\varepsilon^{(l)}$ massimo indipendentemente dall'input nei livelli $l' < l$ "

ossia: P è negata al livello l quando qualunque istanza di P genererebbe un errore di previsione massimo a quel livello, indipendentemente da ciò che i livelli inferiori segnalano. La negazione è tanto più astratta quanto più alto è il livello l a cui si forma, perché i livelli alti sono progressivamente più slegati dall'input sensoriale immediato.

Per la negazione logica piena — quella che opera su proposizioni matematiche o definizioni concettuali — il livello rilevante è quello in cui il modello generativo rappresenta le strutture formali: l'architettura delle dipendenze logiche tra concetti, che è indipendente dall'esperienza sensoriale nel senso in cui la geometria euclidea è indipendente dalla misurazione fisica.

VIII. Il correlato neurale della negazione: evidenze e lacune

Le evidenze sperimentali sull'elaborazione neurale della negazione convergono su tre risultati principali, ciascuno dei quali richiede di essere interpretato nel framework teorico sviluppato.

Il primo risultato è la **negatività frontale tardiva** (*Late Frontal Negativity*, LFN) associata all'elaborazione delle frasi negative. Studi di ERP mostrano che le frasi negative — rispetto alle corrispondenti affermative, controllate per lunghezza e frequenza lessicale — producono una componente negativa nelle regioni frontali, con latenza di 300–500 millisecondi, che riflette il maggiore costo computazionale della negazione. Questa componente è distinta dalla N400 (associata all'anomalia semantica) e dalla P600 (associata alla violazione sintattica): è specifica per la negazione come operazione cognitiva, non per le sue manifestazioni linguistiche locali.

L'interpretazione di questa LFN nel framework proposto è la seguente: la negatività frontale riflette l'attivazione delle regioni prefrontali che devono *costruire e poi inibire* la rappresentazione del contenuto proposizionale prima di produrre la negazione. La latenza tardiva (300–500 ms) è compatibile con la struttura in due fasi — costruzione della rappresentazione affermativa, poi sua modificazione negativa — che la fenomenologia husserliana aveva descritto in termini di stratificazione degli atti intenzionali.

Il secondo risultato, più recente e teoricamente più interessante, riguarda il **Default Mode Network** (DMN) e la sua attivazione durante l'elaborazione della negazione. Studi fMRI mostrano che la comprensione di frasi controfattuali — frasi del tipo "*se Mario fosse arrivato in orario, saremmo partiti*" che descrivono scenari non realizzati — attiva in modo prominente le regioni della DMN: corteccia prefrontale mediale, corteccia cingolata posteriore, ippocampo. Queste regioni sono le stesse attivate durante la simulazione di eventi futuri, la rievocazione autobiografica, e il pensiero ipotetico in generale. L'implicazione è che la negazione controfattuale non è elaborata soltanto come operazione logica sulla proposizione, ma come simulazione dello scenario negato: il cervello *costruisce* il mondo alternativo in cui la proposizione negata è vera, per poi marcare quel mondo come non-attuale.

Questa evidenza è particolarmente significativa perché suggerisce che la negazione proposizionale nel cervello umano non sia implementata come un singolo operatore che inverte il valore di verità, ma come un processo in due fasi che include la simulazione del contenuto negato. In termini del framework predittivo, questo corrisponde alla costruzione di un modello controfattuale — una configurazione del modello generativo che produce previsioni incompatibili con il modello attuale del mondo — seguita dalla sua marcatura come appartenente a uno scenario non-attuale. La negazione non cancella la rappresentazione: la *contestualizza* in un mondo impossibile o non-realizzato.

Il terzo risultato riguarda la **disabilitazione dell'azione** associata alla negazione. Studi di tempi di reazione mostrano che le frasi negative producono una interferenza sistematica con l'esecuzione di azioni compatibili con il contenuto della proposizione negata: leggere "*non afferrare la matita*" rallenta i tempi di risposta per azioni di afferrare rispetto alle azioni di non-afferrare, anche quando la risposta è compatibile con il contenuto della negazione. Questo *effetto di compatibilità negativa* suggerisce che il cervello elabori il contenuto proposizionale della frase — inclusa la sua componente motoria, attivando i sistemi di pianificazione dell'azione — *prima* di elaborare la negazione che lo cancella.

Questo risultato è il correlato comportamentale della struttura husserliana della posizione-e-cancellazione: il cervello afferma prima di negare, attiva prima di inibire. La negazione è un processo di secondo ordine che opera su rappresentazioni già costruite — non un semplice bit-flip, ma una trasformazione di uno stato cognitivo già pienamente formato.

IX. La negazione e il suo opposto: asimmetria tra affermare e negare nel cervello

Uno degli aspetti più sorprendenti della negazione cerebrale è la **profonda asimmetria** tra l'elaborazione di frasi affermative e quella di frasi negative — un'asimmetria che va ben al di là del maggiore costo computazionale già menzionato, e che tocca la struttura stessa delle rappresentazioni prodotte.

Quando si elabora la frase "*la mela è rossa*", la rappresentazione prodotta — secondo le evidenze di priming, tempi di reazione, e neuroimaging — è strutturata come una simulazione percettiva: si attivano le aree visive che processano il colore rosso, le aree motorie associate all'afferramento di oggetti, la corteccia olfattiva e gustativa associata alla mela. La comprensione linguistica è *embodied* nel senso di Barsalou: radica le rappresentazioni linguistiche in simulazioni percettivo-motorie distribuite nelle cortecce sensoriali e motorie.

Quando si elabora la frase "*la mela non è rossa*", la situazione è molto meno chiara. Alcune evidenze suggeriscono che la rappresentazione della mela-rossa venga costruita e poi parzialmente soppressa; altre evidenze suggeriscono che venga costruita e mantenuta attiva insieme alla sua marcatura negativa. La domanda aperta è se la negazione *abolisca* la simulazione percettiva del contenuto negato o la *mantenga attiva con un tag* che la contrassegna come non-attuale.

La risposta, nei dati disponibili, sembra dipendere dal contesto e dalla profondità dell'elaborazione. In compiti di comprensione superficiale e rapida, la simulazione percettiva del contenuto negato viene soppressa — il sistema produce una rappresentazione della non-mela-rossa che è più vicina alla mela-verde che alla mela-rossa. In compiti che richiedono elaborazione profonda e ragionamento sul contenuto, la simulazione percettiva del contenuto negato rimane attiva e accessibile, come se il sistema mantenesse due rappresentazioni parallele: il mondo attuale (mela non rossa) e il mondo controfattuale (mela rossa) implicato dalla negazione.

Questa duplicità — negazione come soppressione vs. negazione come dual-coding — corrisponde alla distinzione tra due funzioni cognitive diverse della negazione. La negazione descrittiva (*la mela non è rossa* come descrizione dello stato del mondo) tende verso la soppressione: il sistema aggiorna il proprio modello del mondo cancellando la proprietà negata. La negazione inferenziale (*se la mela non fosse rossa, allora...*) tende verso il dual-coding: il sistema mantiene attivo il contenuto negato come premessa di un ragionamento ipotetico.

Nel framework della codifica predittiva, questa distinzione corrisponde alla differenza tra aggiornamento del modello (la negazione modifica il prior del sistema, sopprimendo la previsione contraddetta) e mantenimento del modello controfattuale (la negazione costruisce un modello alternativo che viene tenuto attivo in parallelo al modello principale, senza sostituirlo). Il primo tipo di

negazione cambia il paesaggio energetico del sistema; il secondo crea un attrattore temporaneo che coesiste con l'attrattore principale del modello corrente.

Questa coesistenza di attrattori corrispondenti a mondi proposizionali diversi — il mondo reale e il mondo controfattuale implicato dalla negazione — è forse il fatto più sorprendente e meno compreso della cognitiva della negazione. Come il cervello gestisce questa molteplicità di modelli paralleli senza perdere il senso di quale sia attuale, senza confondere il reale e il controfattuale, è una delle domande più profonde che il problema della negazione pone alla neuroscienza. E la risposta, come si argomenterà nella parte successiva, richiede di comprendere non soltanto la meccanica delle rappresentazioni ma la struttura di ciò che Husserl chiamava — con un termine che merita di essere preso sul serio nella sua radicalità — la *posizionalità* degli atti intenzionali: il fatto che ogni atto cognitivo non soltanto rappresenti il proprio oggetto, ma lo situi in un orizzonte di realtà o di irrealtà che determina il suo peso pratico per il soggetto.

L'ontogenesi della negazione e la simulazione ippocampale dell'assente

X. Tre negazioni, tre età: la stratificazione ontogenetica

Se la negazione proposizionale piena è un'operazione cognitiva di secondo ordine — che presuppone la costruzione della rappresentazione affermativa e la sua successiva modificazione — allora ci si aspetta che la sua acquisizione nel corso dello sviluppo non avvenga in un colpo solo, ma attraverso stadi distinti che corrispondono all'emergenza progressiva delle capacità cognitive sottostanti. E in effetti, la letteratura sullo sviluppo del linguaggio documenta con notevole precisione una sequenza di tre stadi nell'acquisizione della negazione che non è né arbitraria né contingente: è, come si argomenterà, la traccia ontogenetica della struttura computazionale della negazione stessa.

Il primo stadio — documentato nei bambini tra i diciotto e i ventiquattro mesi — è la **negazione di rifiuto** o di protesta: il bambino usa la negazione per opporsi a un'azione che un adulto sta compiendo o proponendo. Il *no* del bambino di diciotto mesi quando gli si offre il cibo che non vuole è la forma più primitiva della negazione, e ha una struttura radicalmente diversa dalla negazione logica adulta. Non è la negazione di una proposizione: è la negazione di un'azione in corso, di un evento nel mondo. Ha una funzione performativa e pragmatica — è un atto di rifiuto, non un giudizio — e non richiede la costruzione di nessuna rappresentazione proposizionale. Neurologicamente, è correlata con i sistemi limbici e dell'amigdala che regolano le risposte di approccio-evitamento: è, in senso proprio, negazione come inibizione comportamentale, la forma più diretta in cui l'architettura neurale dell'inibizione si manifesta nel linguaggio.

Il secondo stadio — tra i ventiquattro e i trentasei mesi — è la **negazione di non-esistenza**: il bambino usa la negazione per segnalare l'assenza di un oggetto o di un agente atteso. *"La palla non c'è"*, *"Mamma non c'è"* sono le forme tipiche di questo stadio. Questa negazione è già più vicina alla negazione proposizionale adulta: ha un contenuto referenziale determinato (la palla, la mamma), una struttura predicativa (il soggetto e il predicato di assenza), e riflette la comparsa della **permanenza dell'oggetto** nel senso piagetiano — la capacità di mantenere la rappresentazione di un oggetto anche in sua assenza. Non a caso, questo stadio coincide approssimativamente con il completamento del processo di mielinizzazione delle connessioni tra ippocampo e corteccia prefrontale — le connessioni che permettono il mantenimento attivo di rappresentazioni in assenza di input.

Ma questa negazione di non-esistenza è ancora incompleta rispetto alla negazione proposizionale piena: afferma l'assenza di un oggetto nel mondo, non la falsità di una proposizione su di esso. *"La palla non c'è"* è un giudizio percettivo, non un giudizio logico: non presuppone la costruzione della proposizione *"la palla c'è"* e la sua successiva negazione, ma registra direttamente una discrepanza tra l'aspettativa (la palla è qui) e la percezione (nessuna palla). È, in termini del framework predittivo, un errore di previsione percettivo comunicato verbalmente — la vocalizzazione dell'errore di previsione visivo.

Il terzo stadio — tra i trentasei e i quarantotto mesi — è la **negazione proposizionale vera**: il bambino diventa capace di negare il contenuto di una proposizione indipendentemente dallo stato percettivo corrente. *"Non è vero che i dinosauri vivono ancora"*, *"la Befana non esiste davvero"*, *"papà non ha detto che posso"*. Queste negazioni operano su rappresentazioni proposizionali astratte — non su oggetti percepiti o su azioni in corso — e richiedono la capacità di costruire una rappresentazione proposizionale e di sottoporla a valutazione negativa. L'emergenza di questa capacità coincide con lo sviluppo della Teoria della Mente nel senso di Wimmer e Perner — la capacità di rappresentare le credenze altrui come potenzialmente diverse dalle proprie, e dunque di distinguere tra la proposizione e il suo valore di verità nel modello di un interlocutore specifico.

Questa sequenza ontogenetica — rifiuto, non-esistenza, falsità proposizionale — non è casuale: è la traccia dello sviluppo cognitivo nella sua direzione di progressiva de-contestualizzazione. Il rifiuto è la negazione più contestualizzata (è sempre qui, ora, rispetto a questa azione); la non-esistenza è già parzialmente de-contestualizzata (riguarda un oggetto assente, non un'azione presente); la falsità proposizionale è completamente de-contestualizzata (riguarda il valore di verità di una proposizione indipendentemente dal contesto percettivo immediato). Il percorso dalla negazione di rifiuto alla negazione proposizionale è il percorso dall'inibizione comportamentale alla logica — e questo percorso ha una durata di circa due anni, occupa il periodo di maggiore plasticità prefrontale, e lascia tracce nella struttura delle connessioni corticali che sono misurabili fino all'età adulta.

XI. L'ippocampo come simulatore del non-accaduto

Il risultato più sorprendente delle neuroscienze cognitive degli ultimi quindici anni — un risultato che ha richiesto di rivedere radicalmente la concezione tradizionale dell'ippocampo come semplice archivio della memoria episodica — è che questa struttura sottocorticale svolge un ruolo cruciale non soltanto nella rievocazione del passato ma nella **simulazione del futuro e del controfattuale**. L'ippocampo, nella visione emergente, è un sistema di ricombinazione flessibile di elementi dell'esperienza in configurazioni nuove che possono rappresentare eventi non ancora accaduti, eventi che non accadranno mai, eventi che avrebbero potuto accadere ma non sono accaduti.

Questo ruolo dell'ippocampo come simulatore del non-attuale è stato messo in luce da una convergenza di risultati sperimentali. Il lavoro di Schacter e Addis — in un articolo del 2007 su *Nature* — ha mostrato che le regioni cerebrali attivate durante la rievocazione episodica e quelle attivate durante la simulazione di eventi futuri sono in larga misura le stesse, con l'ippocampo come nodo centrale in entrambi i casi. Questo risultato ha portato alla formulazione della **Constructive Episodic Simulation Hypothesis**: la funzione primaria della memoria episodica non sarebbe la conservazione fedele del passato, ma la dotazione di elementi esperienziali ricomponibili in simulazioni flessibili di scenari non-presenti — passati, futuri, o puramente ipotetici.

Il meccanismo cellulare attraverso cui l'ippocampo realizza questa simulazione è documentato con crescente precisione. Le **place cells** dell'ippocampo — neuroni che scaricano selettivamente quando l'animale si trova in posizioni specifiche dell'ambiente — mostrano durante il sonno e durante i periodi di quiet wakefulness pattern di riattivazione che non soltanto replicano le sequenze di posizioni esperite in precedenza, ma generano sequenze di posizioni *mai esperite*: percorsi alternativi, scorciatoie non prese, destinazioni non raggiunte. Questo fenomeno — documentato da Pfeiffer e Foster nel ratto, e con metodi meno diretti in soggetti umani attraverso fMRI — è la controparte cellulare della costruzione controfattuale: l'ippocampo genera simulazioni di traiettorie nello spazio che non corrispondono alla traiettoria effettiva, ma che potevano essere prese e non lo sono state.

La connessione con la negazione proposizionale è la seguente. La frase "*Mario non è andato al cinema*" non descrive soltanto l'assenza di un evento: implica la costruzione di uno scenario in cui Mario *avrebbe potuto* andare al cinema — lo scenario che la negazione esclude dall'attualità. Questa costruzione implicita dello scenario controfattuale è ciò che rende la negazione semanticamente ricca e pragmaticamente informativa: negare qualcosa di impossibile ("*Mario non è un numero primo*") è categorialmente diverso dal negare qualcosa di possibile ma non accaduto ("*Mario non è andato al cinema*"), e questa differenza riflette precisamente la differenza tra la costruzione di uno scenario ippocampalmente simulabile e la constatazione di un'impossibilità strutturale che non richiede simulazione.

In termini neurali, la proposta è che la negazione proposizionale matura — quella che corrisponde al terzo stadio ontogenetico — sia realizzata da un circuito che include l'ippocampo come generatore della simulazione controfattuale e la corteccia prefrontale ventromediale come sistema di marcatura della realtà (*reality monitoring*): il dispositivo che distingue le simulazioni ippocampali dall'esperienza attuale, mantenendo l'attualità del modello presente mentre le simulazioni controfattuali vengono costruite e valutate.

XII. Il reality monitoring come operatore di negazione neurale

Il concetto di **reality monitoring** — introdotto da Marcia Johnson nei lavori degli anni Ottanta sulla distinzione tra memorie di eventi reali e memorie di eventi immaginati — cattura una funzione cognitiva che è strutturalmente equivalente all'operatore di posizionalità husserliano: il sistema che marca le rappresentazioni come *reali* o *non-reali*, come appartenenti al mondo attuale o a un mondo controfattuale, a un'esperienza percettiva o a un'elaborazione immaginativa.

Il reality monitoring non è un semplice tag binario apposto alle rappresentazioni: è un processo inferenziale che valuta la provenienza delle rappresentazioni sulla base di caratteristiche distribuite. Le rappresentazioni di origine percettiva tendono a essere più ricche di dettagli sensoriali-percettivi (colore, suono, texture), più precise nelle coordinate spazio-temporali, meno legate a processi di elaborazione esplicita. Le rappresentazioni di origine immaginativa — incluse le simulazioni ippocampali del non-accaduto — tendono a essere più schematiche, meno ricche di dettaglio sensoriale, più connesse con stati cognitivi di deliberazione esplicita. Il reality monitoring valuta queste caratteristiche e assegna a ciascuna rappresentazione un grado di realtà — non un valore binario ma un punteggio continuo sulla dimensione reale-immaginato.

Nel framework della codifica predittiva, il reality monitoring corrisponde alla valutazione della **precisione** (precision weighting) delle rappresentazioni: le rappresentazioni di origine percettiva hanno alta precisione (bassa varianza, forte ancoraggio all'input sensoriale), mentre le rappresentazioni di origine immaginativa hanno bassa precisione (alta varianza, debole ancoraggio all'input). L'operatore di negazione neurale consiste nel costruire una simulazione ippocampale del contenuto proposizionale negato con bassa precisione — marcandola come appartenente allo spazio delle possibilità non-attuali — mentre il modello del mondo attuale mantiene alta precisione nel modello che nega la proposizione simulata.

Formalmente, si può esprimere questa struttura nel linguaggio della codifica predittiva. Sia M il modello del mondo attuale e M' il modello controfattuale costruito durante l'elaborazione della negazione. La negazione proposizionale $\neg P$ è rappresentata come la coesistenza di:

$$\begin{aligned}
q_M(s) &\approx p(s \mid o, \neg P) && [\text{modello attuale con alta precisione: } \Pi_M = \Pi_{\{\text{high}\}}] \\
q_{\{M'\}}(s) &\approx p(s \mid P) && [\text{modello controfattuale con bassa precisione: } \Pi_{\{M'\}} \\
&= \Pi_{\{\text{low}\}}]
\end{aligned}$$

dove Π denota la matrice di precisione (l'inverso della covarianza delle previsioni) e o è l'input sensoriale corrente. Il sistema cognitivo mantiene *entrambe* le distribuzioni, ma le distingue attraverso la loro precisione differenziale: il modello attuale ha previsioni precise che corrispondono all'input; il modello controfattuale ha previsioni imprecise che *non* devono corrispondere all'input — e questa imprecisione è la marcatura neurale del suo status di non-attuale.

La robustezza del reality monitoring — la capacità di mantenere la distinzione tra reale e controfattuale anche sotto stress cognitivo, privazione del sonno, o stati alterati di coscienza — è dunque una funzione della differenziale di precisione tra i due modelli. Quando questo differenziale si riduce — come accade nell'elaborazione superficiale, nel sogno, in certi stati dissociativi — la distinzione tra reale e controfattuale si fa meno netta, e la negazione diventa meno robusta: il sistema può confondere la simulazione del negato con la rappresentazione dell'attuale. Questa confusione ha un nome clinico preciso — la **confabulazione** — ed è uno dei deficit neuropsicologici più istruttivi per la teoria della negazione, come si discuterà nella parte dedicata alla patologia.

XIII. Il ruolo dell'amigdala: negazione emotiva e il caso della soppressione fallita

C'è una forma di negazione che la tradizione logico-filosofica non considera ma che ha una rilevanza neurologica e clinica considerevole: la **negazione emotiva**, il tentativo di sopprimere o negare uno stato affettivo. "*Non ho paura*", "*non sono arrabbiato*", "*non mi importa*": queste negazioni non operano su proposizioni su stati del mondo ma su stati interni del soggetto stesso, e la loro elaborazione neurale è sistematicamente diversa da quella delle negazioni descrittive.

L'amigdala — struttura limbica fondamentale per l'elaborazione delle emozioni, in particolare delle emozioni negative — è coinvolta nella negazione emotiva in un modo che rivela una proprietà paradossale della negazione affettiva: il tentativo deliberato di sopprimere un'emozione negativa ("*non dovrei sentirmi così*", "*devo ignorare questa paura*") produce un'attivazione dell'amigdala maggiore, non minore, rispetto al semplice lasciare che l'emozione si dispieghi. Questo risultato — documentato da Gross e collaboratori negli studi sulla soppressione emotiva e da Ochsner nei paradigmi di rivalutazione cognitiva — è la dimostrazione sperimentale del paradosso della negazione affettiva: tentare di negare uno stato emotivo richiede di attivarlo sufficientemente per poterlo negare, producendo un rimbalzo che amplifica lo stato che si cerca di sopprimere.

Questo paradosso ha una spiegazione formale precisa nel framework della codifica predittiva. La soppressione emotiva è un tentativo di ridurre il segnale di errore di previsione prodotto dall'amigdala — il segnale che comunica la discrepanza tra lo stato emotivo atteso (calma) e lo stato emotivo effettivo (paura). Ma per ridurre questo segnale di errore, il sistema deve *aumentare la precisione* della previsione che lo stato emotivo sia calma — il che richiede un investimento computazionale nelle corteccie prefrontali che paradossalmente amplifica l'attenzione verso il segnale di paura che si cerca di sopprimere. La soppressione emotiva è un caso in cui l'operatore di negazione neurale produce l'effetto opposto a quello inteso, perché opera su un sistema (l'amigdala) che non ha accesso diretto agli stati proposizionali ma risponde alle previsioni di alto livello attraverso connessioni discendenti che ne amplificano, non ne sopprimono, la risposta.

La rivalutazione cognitiva — il reinterpretare la situazione in un modo che ne riduca il carica emotiva — funziona invece molto meglio, come mostrano i dati di Ochsner: non tenta di negare lo stato emotivo, ma modifica il modello generativo dell'evento che lo ha prodotto, in modo che la previsione generata dall'alto sia compatibile con una risposta affettiva meno intensa. Non è una negazione dell'emozione ma una modifica della sua *premessa causale* — e questa distinzione corrisponde esattamente alla differenza, nel framework della logica lineare, tra consumare irreversibilmente la risorsa (modificare il modello generativo) e tentare di invertire il bit senza modificare la struttura che lo produce (soppressione).

Questa asimmetria tra soppressione e rivalutazione è uno degli argomenti empirici più forti a favore della tesi che la negazione biologicamente efficace non sia mai la semplice inversione di un segnale, ma sempre la modifica della struttura generativa che produce il segnale — una forma di negazione che opera sul *modello* piuttosto che sul *risultato*, sulla causa piuttosto che sull'effetto.

XIV. Sviluppo prefrontale e capacità di negazione controfattuale: la finestra critica

L'acquisizione della negazione proposizionale piena — il terzo stadio ontogenetico — coincide con un periodo di rapido sviluppo della corteccia prefrontale, e in particolare delle sue connessioni con l'ippocampo e con la corteccia cingolata anteriore. Questo non è casuale: la capacità di costruire simulazioni controfattuali di alta qualità, di mantenerle distinte dal modello del mondo attuale, e di valutare le proposizioni sulla base di queste simulazioni, dipende dalla maturazione di un circuito specifico che include tre componenti distinte.

La prima componente è la **corteccia prefrontale ventromediale** (vmPFC), coinvolta nel reality monitoring e nella valutazione della realtà delle rappresentazioni. La vmPFC riceve input dall'ippocampo e dall'amigdala, e le sue

connessioni con il sistema limbico la rendono sensibile sia al contenuto proposizionale delle simulazioni sia al loro tono affettivo — la valenza emotiva del mondo controfattuale costruito. La maturazione della vmPFC è un processo lento che si prolunga fino alla tarda adolescenza, il che spiega perché i bambini piccoli abbiano difficoltà con le negazioni proposizionali complesse che implicano ragionamento controfattuale a distanza (se P fosse vero, e se Q fosse vero, allora $\neg R$ — questo tipo di negazione concatenata richiede vmPFC matura).

La seconda componente è la **corteccia prefrontale dorsolaterale** (dlPFC), coinvolta nella memoria di lavoro e nel mantenimento attivo di multiple rappresentazioni simultanee. La capacità di tenere presenti contemporaneamente il modello attuale e il modello controfattuale — senza che il secondo invada e disorganizzi il primo — richiede una memoria di lavoro sufficientemente capiente da contenere entrambi e un meccanismo di separazione che li mantenga distinti. La dlPFC fornisce entrambe queste funzioni, e la sua maturazione — che procede in parallelo con quella della vmPFC, con piena maturità strutturale solo nella tarda adolescenza — è la condizione neurobiologica della negazione proposizionale adulta.

La terza componente, forse la meno ovvia, è la **corteccia cingolata anteriore** (ACC) nella sua divisione cognitiva. L'ACC cognitiva è il sistema di rilevazione dei conflitti: risponde con forte attivazione quando il sistema deve gestire rappresentazioni incompatibili o risposte concorrenti. Durante l'elaborazione della negazione proposizionale, l'ACC deve gestire il conflitto tra la simulazione del contenuto negato (che attiva il modello controfattuale) e il mantenimento del modello attuale (che lo esclude): questo conflitto è strutturale e inevitabile, non un errore computazionale. L'ACC ne monitora la risoluzione, segnalando ai sistemi prefrontali la necessità di aumentare il controllo cognitivo per mantenere la separazione tra reale e controfattuale.

L'interazione di questi tre sistemi — vmPFC per il reality monitoring, dlPFC per il mantenimento parallelo, ACC per il monitoraggio del conflitto — costituisce il circuito neurale della negazione proposizionale matura. La sua maturazione progressiva nel corso dello sviluppo è la base neurologica della sequenza ontogenetica in tre stadi: la maturazione dell'ACC permette il rifiuto comportamentale; la maturazione delle connessioni ippocampo-vmPFC permette la negazione di non-esistenza; la maturazione completa del circuito prefrontale permette la negazione proposizionale piena e il ragionamento controfattuale.

La discontinuità filogenetica e la topologia dell'assenza

XV. Il confine filogenetico: dove la negazione diventa proposizionale

La domanda sulla discontinuità filogenetica della negazione — perché soltanto gli esseri umani sembrano capaci di negazione proposizionale piena — è tanto più acuta quanto più si considera che l'inibizione neurale è presente in *tutti* i vertebrati e in gran parte degli invertebrati, e che molte specie animali mostrano comportamenti che sembrano implicare una forma di negazione cognitiva. Il corvo che nasconde il cibo soltanto quando non è osservato da altri corvi sembra tenere conto dell'assenza di testimoni — una forma primitiva di rappresentazione del non-vedere. Lo scimpanzé che, in certi paradigmi sperimentali, risponde correttamente alle domande su oggetti nascosti sembra rappresentare la loro assenza dalla propria percezione. Eppure nessuna di queste specie produce o comprende negazioni proposizionali nella forma in cui le producono gli esseri umani.

La differenza non riguarda l'inibizione comportamentale — capacità di non fare ciò che si è propensi a fare — che è ampiamente diffusa e che corrisponde al primo stadio ontogenetico della negazione. Non riguarda nemmeno la rappresentazione dell'assenza percettiva — la registrazione che un oggetto atteso non è presente — che sembra accessibile ai primati superiori e ad alcune altre specie. La differenza riguarda la capacità di *operare su proposizioni come oggetti di valutazione*: di prendere una struttura proposizionale, valutarne il valore di verità rispetto a un modello del mondo, e comunicare il risultato di questa valutazione attraverso un enunciato esplicito.

Questa capacità richiede almeno tre ingredienti che sembrano esclusivamente o prevalentemente umani nella loro forma piena. Il primo è la **ricorsività proposizionale**: la capacità di annidare proposizioni dentro proposizioni, di negare la negazione, di costruire strutture come "*non è vero che non è vero che P*" che richiedono di operare su proposizioni come oggetti di secondo ordine. La ricorsività, come si è argomentato nel saggio sulle oscillazioni, è la proprietà distintiva della sintassi umana — e la negazione proposizionale la richiede nella sua forma piena.

Il secondo ingrediente è la **Teoria della Mente di secondo ordine**: la capacità di rappresentare non soltanto le credenze altrui ma le credenze altrui *sulla proprie credenze*. La negazione proposizionale comunicativa — dire "*non è così*" a un interlocutore — non ha senso senza la capacità di rappresentare che l'interlocutore creda qualcosa di diverso da ciò che si crede, e di mirare esplicitamente alla modifica di questa credenza. I primati non umani mostrano forme rudimentali di Teoria della Mente di primo ordine, ma le evidenze per la Teoria della Mente di secondo ordine — necessaria per la negazione proposizionale pienamente comunicativa — rimangono molto più deboli e controverse.

Il terzo ingrediente è la **simulazione controfattuale generativa**: la capacità ippocampale di costruire scenari non-attuali di qualità sufficientemente alta da servire come base per la valutazione proposizionale. Come si è argomentato nella sezione precedente, la negazione proposizionale adulta richiede la costruzione

implicita dello scenario che viene negato — e questa costruzione è tanto più efficace quanto più ricca e dettagliata è la simulazione ippocampale. La ricchezza delle simulazioni ippocampali è correlata con la densità e la complessità delle connessioni ippocampo-corticali, che nel cervello umano raggiungono un grado di sviluppo senza eguali nei primati non umani.

La convergenza di questi tre ingredienti — ricorsività sintattica, Teoria della Mente di secondo ordine, simulazione controfattuale generativa — nel cervello umano produce la capacità di negazione proposizionale piena. Non è un singolo meccanismo evolutivamente nuovo: è l'emergenza di una proprietà sistemica da una combinazione di capacità che, prese isolatamente, hanno precursori in altre specie. La negazione proposizionale è, come il linguaggio ricorsivo di cui è parte integrante, una proprietà emergente — una proprietà del sistema che non è riducibile a nessuno dei suoi componenti presi separatamente.

XVI. La teoria delle categorie e la negazione: topos, sottooggetti e classificatori di verità

C'è un territorio matematico che il dibattito sulla negazione non ha mai sistematicamente esplorato, benché offra forse il quadro formale più generale e strutturalmente illuminante per pensare la negazione in sistemi cognitivi: la **teoria delle categorie**, e in particolare la teoria dei **topos**. Portare la teoria dei topos alla neuroscienza della negazione è un salto disciplinare audace — e va dichiarato apertamente che ciò che segue ha lo status di proposta formale speculativa, non di teoria verificata. Ma è speculazione rigorosa, nel senso che le strutture matematiche che si introducono hanno proprietà precise che producono previsioni precise sulle proprietà che la negazione neurale deve avere.

Un **topos** è, nella definizione tecnica, una categoria cartesianamente chiusa con un classificatore di sottooggetti. Per i fini di questo saggio, è sufficiente caratterizzarlo attraverso le sue proprietà rilevanti. Un topos generalizza la struttura degli insiemi — è una categoria in cui si può fare inferenza logica, costruire prodotti e coprodotti, definire funzioni esponenziali — ma in modo che la logica interna del topos non sia necessariamente quella booleana classica. In un topos generico, i valori di verità non sono soltanto $\{0, 1\}$ ma formano un oggetto ricco — l'**oggetto classificatore di verità** Ω — che può avere una struttura molto più complessa.

L'oggetto Ω di un topos è definito dalla proprietà universale seguente: per ogni sottooggetto S di un oggetto A (ossia per ogni "parte" di A), esiste un'unica freccia caratteristica $\chi_S: A \rightarrow \Omega$ che "classifica" S — che è vero (nel senso del topos) esattamente sugli elementi di S e falso sugli altri. In un topos di insiemi standard, $\Omega = \{0, 1\}$ e le frecce caratteristiche sono le funzioni indicatrici booleane. In un topos più generale, Ω può avere molti più elementi — e dunque la logica del topos ammette più valori di verità.

La negazione nel contesto di un topos è definita come un morfismo $\neg: \Omega \rightarrow \Omega$ — una trasformazione dell'oggetto classificatore di verità su se stesso. Nella logica booleana, questo morfismo è la semplice inversione: $\neg(1) = 0$, $\neg(0) = 1$. Ma in un topos intuizionista — come il topos dei fasci (*sheaves*) su uno spazio topologico — il morfismo di negazione ha una struttura molto più ricca. Per un aperto U dello spazio base, la negazione di una sezione locale è:

$$\neg P(U) = \text{int}(X \setminus \text{cl}(P(U)))$$

dove int denota l'interno e cl la chiusura nella topologia dello spazio base. Questa formula dice: la negazione di P sull'aperto U è l'aperto più grande che non si sovrappone con la chiusura dell'insieme su cui P è vero. Non è la semplice complementazione, ma l'*interno del complemento della chiusura* — una doppia operazione topologica che corrisponde alla negazione intuizionista robusta: $\neg P$ è vera soltanto dove P è non soltanto attualmente falsa ma strutturalmente irraggiungibile.

La struttura topologica di questa negazione è visivamente chiara: l'interno del complemento della chiusura è sempre contenuto nel complemento, ma può essere strettamente più piccolo — lascia una "zona di confine" in cui né P né $\neg P$ sono definite, una regione di indecidibilità che corrisponde esattamente alla zona epistemica di incertezza genuina che la logica intuizionista ammette e quella classica non può rappresentare.

XVII. Il classificatore di verità neurale: precisione come valore di verità continuo

La struttura matematica del topos suggerisce una proposta neurale specifica: il cervello non opera con un classificatore di verità booleano $\{0, 1\}$ ma con un classificatore di verità **continuo** — la precisione delle rappresentazioni, già introdotta nel framework della codifica predittiva. La precisione Π di una rappresentazione è il reciproco della sua varianza: $\Pi = \Sigma^{-1}$. Rappresentazioni con alta precisione sono "più vere" nel senso in cui il sistema vi si affida per guidare l'azione; rappresentazioni con bassa precisione sono "meno vere" — mantenute attive ma con peso ridotto nelle inferenze e nelle decisioni.

In questa prospettiva, il cervello opera con un oggetto classificatore di verità Ω_{neurale} che è non l'insieme discreto $\{0, 1\}$ ma l'intervallo continuo $[0, \infty)$ dei possibili valori di precisione — o, dopo una trasformazione monotona, l'intervallo $[0, 1]$ delle probabilità soggettive. La logica neurale è dunque, strutturalmente, una logica a valori continui — il che corrisponde al fatto che il cervello non afferma mai proposizioni con certezza assoluta ma sempre con gradi di credenza che riflettono la quantità e la qualità delle evidenze disponibili.

La negazione nel classificatore di verità neurale continuo non è la semplice inversione $\Pi \rightarrow -\Pi$ (che non ha senso per valori di precisione, che sono sempre positivi), ma qualcosa di più complesso. Si propone la seguente formalizzazione. Sia $p(P) \in [0, 1]$ la credenza del sistema nella proposizione P — la sua probabilità soggettiva. La credenza nella negazione $\neg P$ è:

$$p(\neg P) = 1 - p(P) \quad [\text{negazione classica}]$$

nella logica booleana a valori continui (probabilità). Ma questa forma di negazione non distingue tra la negazione *epistemica* (non so se P , quindi assegno $p(P) = 0.5$ e $p(\neg P) = 0.5$) e la negazione *proposizionale* (so che $\neg P$, quindi $p(P) = 0$ e $p(\neg P) = 1$). Per distinguere questi casi, è necessario considerare non soltanto il valore di $p(P)$ ma la struttura della distribuzione di probabilità — in particolare, la sua curtosi e la sua entropia.

Una distribuzione piatta (alta entropia, bassa curtosi) su P e $\neg P$ corrisponde all'ignoranza epistemica — l'equivalente neurale della negazione debole della semantica di Kripke. Una distribuzione concentrata su $\neg P$ (bassa entropia, alta curtosi nella regione di $\neg P$) corrisponde alla negazione proposizionale piena — l'equivalente neurale della negazione forte. La transizione tra queste due strutture non è discreta ma continua, e può essere misurata attraverso l'entropia della distribuzione di credenza:

$$H(P) = -p(P) \cdot \log p(P) - (1 - p(P)) \cdot \log(1 - p(P))$$

con $H(P)$ massima ($= 1$ bit) per $p(P) = 0.5$ (ignoranza completa) e $H(P) = 0$ per $p(P) = 0$ o $p(P) = 1$ (certezza nella negazione o nell'affermazione). La negazione proposizionale neurale è dunque caratterizzata da *bassa entropia* nella distribuzione di credenza — e questa proprietà è misurabile, in via indiretta, attraverso le fluttuazioni del segnale BOLD nelle aree prefrontali che implementano il reality monitoring.

XVIII. Il morfismo di negazione come funzione di connettività: una proposta computazionale

La formalizzazione del classificatore di verità neurale come oggetto continuo $[0,1]$ permette di formulare una proposta computazionale esplicita sul morfismo di negazione nel cervello. Il morfismo di negazione $\neg: \Omega_{\text{neurale}} \rightarrow \Omega_{\text{neurale}}$ è una funzione che trasforma i valori di precisione delle rappresentazioni in valori di precisione delle loro negazioni. Si propone che questa funzione abbia la forma:

$$\Pi(\neg P) = f(\Pi(P), \Pi(P \rightarrow \perp), \Pi(M))$$

dove $\Pi(P)$ è la precisione della rappresentazione di P , $\Pi(P \rightarrow \perp)$ è la precisione della rappresentazione della derivazione di assurdità da P (il percorso di prova nel senso intuizionista), e $\Pi(M)$ è la precisione del modello generativo M rispetto al quale P viene valutata. Questa formula cattura la struttura stratificata della negazione: la negazione di P non è semplicemente $1 - p(P)$, ma dipende da quanto bene il sistema ha costruito la sua prova di negazione (il termine $P \rightarrow \perp$) e da quanto affidabile è il modello rispetto al quale questa prova è costruita (il termine M).

La funzione f deve soddisfare alcune condizioni di coerenza:

$$\begin{aligned} f(\Pi, \Pi_{\text{proof}}, \Pi_M) &\rightarrow 1 \text{ quando } \Pi_{\text{proof}} \rightarrow \infty \text{ e } \Pi_M \rightarrow \infty \\ f(\Pi, 0, \Pi_M) &= \Pi_M / (\Pi + \Pi_M) \quad (\text{caso di prova assente: credenza a priori del modello}) \\ f(0, \Pi_{\text{proof}}, \Pi_M) &\rightarrow \Pi_{\text{proof}} \quad (\text{caso di } P \text{ completamente incerta}) \end{aligned}$$

Una forma esplicita che soddisfa queste condizioni è:

$$\Pi(\neg P) = \Pi_M \cdot (1 + \Pi_{\text{proof}}/\Pi_M) / (1 + \Pi/\Pi_M) = (\Pi_M + \Pi_{\text{proof}}) / (1 + \Pi/\Pi_M)$$

Questa formula ha la struttura di un aggiornamento bayesiano: la precisione della negazione è la precisione del modello amplificata dalla precisione della prova di negazione e ridotta dalla precisione della proposizione originale. Quando la proposizione P è molto precisa (Π grande), la sua negazione è meno precisa — coerentemente con il fatto che è più difficile negare ciò che si conosce bene; quando la prova di negazione è molto precisa (Π_{proof} grande), la negazione è molto precisa — coerentemente con il fatto che una prova rigorosa produce una negazione robusta.

Questa proposta è verificabile in via indiretta: predice che la precisione delle negazioni proposizionali — misurabile attraverso le fluttuazioni del segnale neurale che codifica la credenza $\neg P$ — sia correlata sia con la precisione della prova di negazione (la qualità del ragionamento che porta alla negazione) sia con la precisione del modello generativo (la qualità complessiva del modello del mondo del soggetto). Soggetti con modelli generativi più ricchi e più precisi dovrebbero produrre negazioni più robuste — una previsione che corrisponde, nella letteratura sulla psicologia del ragionamento, all'osservazione che la qualità del ragionamento sillogistico e controfattuale è correlata con la ricchezza della conoscenza di sfondo.

XIX. La logica paraconsistente e il cervello che tollera la contraddizione

La proposta del classificatore di verità continuo e del morfismo di negazione bayesiano ha una conseguenza che, a prima vista, può sembrare indesiderata: ammette stati in cui sia $p(P)$ che $p(\neg P)$ sono significativamente diverse da zero — stati in cui il sistema crede contemporaneamente a P e a $\neg P$ con credenze non trascurabili. Nella logica classica, questo è una contraddizione che porta all'esplosione (*ex contradictione quodlibet*: da una contraddizione segue qualunque proposizione). Ma il cervello umano è pieno di questi stati — stati di credenza ambivalente, di dissonanza cognitiva, di atteggiamenti contrastanti verso lo stesso oggetto — e non esplode.

La disciplina logica che studia i sistemi formali tolleranti alla contraddizione è la **logica paraconsistente**, sviluppata indipendentemente da Newton da Costa e da Graham Priest. Un sistema paraconsistente è un sistema in cui la regola *ex contradictione quodlibet* non vale: una contraddizione locale non si propaga a tutto il sistema. La logica della rilevanza (*relevance logic*) e la logica dialeteista di Priest, che ammette la verità di proposizioni della forma $P \wedge \neg P$ in certi contesti, sono i principali sistemi paraconsistenti sviluppati in letteratura.

Per il cervello, la paraconsistenza non è una patologia ma una caratteristica adattiva. Un sistema che, di fronte a evidenze contrastanti, aggiornasse istantaneamente le proprie credenze verso la certezza assoluta in una direzione o nell'altra sarebbe cognitivamente rigido e vulnerabile agli errori sistematici. Un sistema che mantiene credenze contrastanti con pesi proporzionali alle evidenze è più flessibile, più capace di accumulare evidenze graduali, e più robusto rispetto al rumore nelle singole osservazioni. La paraconsistenza neurale — la capacità di mantenere $p(P)$ e $p(\neg P)$ entrambi non trascurabili per periodi prolungati — è la condizione di possibilità della formazione graduale di credenze solide in un ambiente informativo rumoroso.

Il correlato neurale di questa paraconsistenza è la **coesistenza di attrattori** nello spazio degli stati del sistema cognitivo: il modello che afferma P e il modello che afferma $\neg P$ coesistono come attrattori distinti ma entrambi accessibili, con bacini di attrazione che si sovrappongono parzialmente nella regione di credenza ambivalente. La transizione tra i due attrattori — il processo di formazione di una credenza definitiva — è una biforcazione dinamica in cui il sistema sceglie (deterministicamente o stocasticamente) quale attrattore raggiungere sulla base delle evidenze accumulate. La negazione proposizionale piena corrisponde al raggiungimento dell'attrattore $\neg P$ con un bacino di attrazione sufficientemente profondo da rendere il ritorno all'attrattore P non spontaneo.

Questa struttura ha una conseguenza empirica precisa per la psicologia del ragionamento: la robustezza della negazione — la resistenza dell'agente a essere convinto del contrario — non è una proprietà della proposizione negata ma una proprietà del paesaggio energetico del sistema in cui la negazione è stata formata. Una negazione formata attraverso un lungo processo di accumulazione di evidenze convergenti ha un bacino di attrazione profondo e un'alta barriera di potenziale verso l'attrattore opposto; una negazione formata su evidenze scarse

ha un bacino poco profondo e una bassa barriera. Misurare la profondità del bacino — attraverso paradigmi di resistenza alla persuasione, di aggiornamento bayesiano, o di tempi di risposta in compiti di rievocazione — è misurare la robustezza della negazione, e dunque la qualità della prova su cui essa si fonda.

La patologia della negazione e le sue implicazioni filosofiche

XX. Quando la negazione si rompe: confabulazione, anosognosia e la perdita del reality monitoring

Le patologie della negazione offrono, come sempre in neuropsicologia, una finestra privilegiata sulla struttura del meccanismo sano. La perdita selettiva della capacità di negare — o, simmetricamente, la produzione sistematica di negazioni false — rivela i componenti del sistema e le loro relazioni, nel modo in cui le lesioni anatomiche rivelano la funzione dei circuiti che distruggono. Due sindromi meritano un'attenzione particolare, non tanto per la loro frequenza clinica quanto per la loro rilevanza teorica: la confabulazione e l'anosognosia.

La **confabulazione** — la produzione inconsapevole di narrazioni false su eventi non accaduti, senza intenzione di ingannare e senza consapevolezza della loro falsità — è una perturbazione del reality monitoring che illustra con drammatica chiarezza la dissociabilità tra la costruzione di simulazioni e la loro marcatura come controfattuali. Il paziente confabulante costruisce simulazioni ippocampali di eventi non accaduti — ricordi di conversazioni mai avute, di azioni mai compiute, di luoghi mai visitati — e le presenta come memorie reali, perché il sistema che dovrebbe marcarle come simulazioni (la vmPFC e il suo circuito) è danneggiato o disconnesso.

Nel framework proposto, la confabulazione è una perturbazione del morfismo di negazione neurale: il sistema può costruire rappresentazioni del non-accaduto (la simulazione ippocampale funziona) ma non può assegnare loro il differenziale di precisione che le distingue dal passato reale (il reality monitoring è compromesso). Formalmente, la funzione f che determina $\Pi(\neg P)$ perde il suo termine di dipendenza da $\Pi(M)$ — il modello generativo non contribuisce più alla valutazione della credibilità della simulazione — e il risultato è che la simulazione del non-accaduto riceve la stessa precisione della memoria del passato reale: le due distribuzioni di credenza collassano verso la stessa forma, eliminando la distinzione tra reale e controfattuale.

Questo non è soltanto il fallimento della negazione di proposizioni specifiche: è il collasso della struttura che rende possibile la negazione in quanto tale. Il paziente confabulante non è incapace di usare la parola *no* — può rifiutare cibi, può protestare, può negare eventi singoli quando interrogato direttamente. Ma è incapace di mantenere la struttura di fondo che separa il mondo reale dai mondi

possibili non-attuali — la struttura che Husserl chiamava *posizionalità* e che qui si è proposto di identificare con il reality monitoring prefrontale. Senza questa struttura, la negazione proposizionale perde il suo senso: non si nega rispetto a nessun mondo, perché nessun mondo ha la priorità degli altri.

L'anosognosia — la mancanza di consapevolezza di una propria disabilità, tipicamente osservata in pazienti con paralisi degli arti dopo lesioni dell'emisfero destro — è un caso ancora più radicale di perturbazione della negazione. Il paziente anosognosico non nega la disabilità nel senso di affermare "*non sono paralizzato*": nega nel senso profondo di non poter costruire la simulazione del proprio stato attuale come oggetto di valutazione. La paralisi non entra nel modello generativo del paziente perché il sistema che dovrebbe rilevare la discrepanza tra il comportamento motor atteso e il feedback propriocettivo assente — tipicamente circuiti nell'emisfero destro — è distrutto dalla lesione.

In termini del framework predittivo, l'anosognosia è una perturbazione del meccanismo di errore di previsione al livello del modello corporeo: il segnale gamma che dovrebbe segnalare la discrepanza tra la previsione di movimento e l'assenza di movimento effettivo non viene generato — non perché il sistema neghi consapevolmente la paralisi, ma perché il dispositivo che produce la negazione biologica primaria (l'errore di previsione) è danneggiato. Il paziente non afferma "*il mio braccio si muove*": semplicemente non nega questa proposizione, perché non ha accesso al segnale di errore che la confuterebbe. È la forma più pura dell'assenza di negazione — non la negazione della negazione, ma il silenzio strutturale del sistema che dovrebbe produrla.

XXI. La negazione in psicoanalisi: rimozione, scotomizzazione e la negazione come difesa

C'è una tradizione che ha pensato la negazione in modo sistematico e con conseguenze cliniche rilevanti, e che la neuropsicologia cognitiva ha a lungo ignorato: la tradizione psicoanalitica. Nel breve saggio *Die Verneinung* (1925) — non il saggio di Frege del 1919 che porta lo stesso titolo, ma il testo freudiano che merita di essere distinto — Sigmund Freud aveva proposto una teoria della negazione che anticipa in modo sorprendente alcune delle strutture qui discusse.

Per Freud, la negazione linguistica è il meccanismo attraverso cui il rimosso riesce a emergere nella coscienza mantenendo la sua rimozione: il paziente che dice "*non stavo pensando a mia madre*" porta alla coscienza il pensiero materno proprio attraverso la sua negazione. La negazione è, in questa lettura, non l'assenza del contenuto ma il modo in cui il contenuto si presenta alla coscienza mentre rimane *soggettivamente* escluso — la sua presenza nella forma dell'esclusione.

Questa intuizione freudiana ha una struttura formalmente interessante che il framework teorico qui sviluppato permette di articolare con precisione. La

negazione freudiana corrisponde a un caso particolare del dual-coding discusso nella sezione precedente: il contenuto rimosso viene mantenuto attivo come simulazione controfattuale (è presente come il *non-io* del sistema cognitivo) ma non viene integrato nel modello del mondo attuale (rimane marcato come escluso dalla posizionalità husserliana). La negazione linguistica che il paziente produce è la vocalizzazione di questa struttura di esclusione attiva: non la cancellazione del contenuto, ma la sua presentazione nella modalità del rifiuto — che è, non coincidentemente, il primo stadio ontogenetico della negazione.

La **scotomizzazione** — la forma più radicale di difesa psicoanalitica, in cui il soggetto non soltanto rimuove ma non registra nemmeno la percezione del materiale ansiogeno — è strutturalmente diversa dalla rimozione con negazione. Nella scotomizzazione, il segnale di errore di previsione viene soppresso *prima* che raggiunga il livello della coscienza: la simulazione controfattuale non viene nemmeno costruita, perché il sistema ha imparato a non costruirla. È la forma di negazione più profonda: non la negazione di una proposizione già formata, ma la negazione della formazione della proposizione stessa — la struttura che in logica corrisponde alla *negazione dell'antecedente* piuttosto che alla negazione della conclusione.

Questa proposta — che le difese psicoanalitiche siano diverse modi di implementare la negazione a livelli diversi della gerarchia predittiva — non è soltanto un'operazione di traduzione disciplinare. Suggerisce previsioni empiriche testabili: la rimozione con negazione dovrebbe produrre attivazione ippocampale (la simulazione viene costruita) e attivazione vmPFC (il reality monitoring marca la simulazione come esclusa), con inibizione dell'integrazione nel modello corrente; la scotomizzazione dovrebbe produrre inibizione a livello dell'amigdala e delle prime stazioni di processing sensoriale, senza attivazione ippocampale (la simulazione non viene costruita). Paradigmi di neuroimaging su popolazioni con diversi stili difensivi misurabili attraverso strumenti psicometrici validati — la Defence Mechanism Rating Scale o simili — potrebbero testare questa dissociazione.

XXII. Negazione, tempo e la struttura dell'anticipazione

C'è un aspetto della negazione che tutti i framework discussi hanno trattato soltanto marginalmente: la sua relazione con il **tempo** — non il tempo del processamento neurale, ma il tempo del contenuto proposizionale. Le proposizioni si riferiscono a stati del mondo in tempi diversi: passati, presenti, futuri, atemporal. La negazione di una proposizione passata ("*ieri non è piovuto*"), di una proposizione presente ("*non sta piovendo*"), e di una proposizione futura ("*domani non pioverà*") sono cognitivamente diverse, e questa differenza ha correlati neurali che illuminano la struttura del meccanismo di negazione.

La negazione del passato è una negazione della *memoria*: richiede che il modello del passato venga aggiornato per escludere l'evento negato. Questa operazione

coinvolge i processi di riconsolidamento della memoria — il processo attraverso cui le memorie, ogni volta che vengono richiamate, entrano in uno stato di plasticità transitoria in cui possono essere modificate prima di essere nuovamente consolidate. Negare un evento passato, in questo senso, non è soltanto affermare la sua non-accadenza: è una potenziale operazione di modifica del modello mnemonico, che richiede il riconsolidamento attivo della memoria dell'assenza dell'evento.

La negazione del futuro è una negazione dell'*anticipazione*: richiede che la simulazione di un evento futuro atteso venga soppressa o marcata come non-attuale. Nel framework della codifica predittiva, le protenzioni — le aspettative sul futuro immediato — sono parte integrante della gerarchia predittiva, e la loro negazione richiede un aggiornamento delle previsioni che può essere più o meno traumatico a seconda di quanto profondamente le aspettative erano radicate nel modello generativo. La negazione di un evento fortemente atteso — il contraccolpo delle aspettative frustrate — produce un segnale di errore di previsione di grande ampiezza, e questo segnale ha correlati neurali e comportamentali ben documentati: la risposta di delusione, la riorganizzazione delle priorità, la modificazione del modello generativo.

La negazione del presente — la più semplice logicamente, la più immediata fenomenologicamente — è in realtà la più complessa neurologicamente, perché richiede la coesistenza di un input sensoriale attivo (il presente si impone alla percezione) con la sua negazione proposizionale (che costruisce e mantiene il mondo controfattuale in cui il negato è vero). È la struttura del dual-coding nella sua forma più acuta: input e simulazione controfattuale coesistono con la massima contemporaneità, e il reality monitoring deve lavorare al massimo della sua capacità per mantenerli distinti.

Questa analisi suggerisce che il costo computazionale della negazione non sia costante ma vari sistematicamente con la distanza temporale del contenuto negato: le negazioni del futuro prossimo dovrebbero essere le più costose (le aspettative immediate sono le più radicate e le più resistenti alla modifica), seguite dalle negazioni del presente, e dalle negazioni del passato remoto (le memorie remote sono meno plastiche e più facili da marcare come non-attuali senza disturbare il modello corrente). Questa previsione è verificabile attraverso misure di latenza e di ampiezza delle componenti ERP in paradigmi che variano sistematicamente la distanza temporale del contenuto negato.

XXIII. La negazione come condizione di possibilità del pensiero critico

Si può ora tornare al problema filosofico di fondo con il quale si è aperto il saggio, ma con strumenti teorici incomparabilmente più ricchi. La domanda — come può un sistema fondamentalmente attivazionale rappresentare l'assenza come contenuto — non ha ricevuto una risposta univoca, ma ha ricevuto qualcosa di più prezioso: una mappa della sua complessità.

La risposta parziale che emerge dall'analisi è la seguente: il cervello non rappresenta la negazione come un'operazione su valori di verità, ma come una struttura dinamica a più livelli che include la costruzione della rappresentazione affermativa, la sua simulazione controfattuale, il suo marcatura come non-attuale attraverso il differenziale di precisione, e la sua integrazione nella gerarchia predicativa come vincolo sul modello generativo. La negazione non è un bit-flip: è una trasformazione dell'intero stato cognitivo del sistema, che modifica contemporaneamente le rappresentazioni proposizionali, il modello generativo, il paesaggio degli attrattori, e le aspettative sull'input futuro.

Questa complessità non è un difetto del sistema biologico rispetto all'eleganza del sistema logico formale: è una ricchezza funzionale che permette al cervello di fare con la negazione molto più di quanto un sistema logico possa fare. Il cervello non soltanto nega: valuta la robustezza della negazione, ne misura il costo epistemico, mantiene aperta la possibilità di revisione, distingue la negazione epistemica da quella ontologica, usa la negazione come strumento di simulazione del possibile e del controfattuale, e — nei casi più sofisticati di ragionamento ipotetico — costruisce catene di negazioni annidate che producono conclusioni impossibili da ottenere attraverso l'affermazione diretta.

È questa ricchezza della negazione biologica che rende possibile il **pensiero critico** nel suo senso più pieno: non la semplice capacità di dire *no*, ma la capacità di costruire mentalmente un mondo in cui le cose stanno diversamente da come stanno, di valutare quel mondo come alternativo rispetto all'attuale, di usare questa valutazione per testare ipotesi, smontare argomentazioni, riconoscere errori, e aggiornare le proprie credenze in modo coerente con le evidenze. Il pensiero critico è, nella sua struttura più profonda, la negazione in azione — la capacità di opporre al mondo così com'è la simulazione del mondo come potrebbe essere, e di usare questa opposizione per orientarsi meglio nel mondo che effettivamente è.

XXIV. Il paradosso residuo: l'assenza che si presenta

Si è attraversato un lungo percorso — dalla logica intuizionista alla semantica di Kripke, dalla confabulazione alla logica lineare di Girard, dal reality monitoring all'ontogenesi in tre stadi, dalla logica dei topos alla paraconsistenza neurale — e tuttavia il paradosso iniziale non è del tutto dissolto. Rimane qualcosa che resiste, non alla comprensione scientifica ma alla comprensione *del perché* la comprensione scientifica non sia sufficiente.

Il residuo è questo: tutta la spiegazione neurale e formale della negazione è una spiegazione del *meccanismo* attraverso cui il cervello produce e gestisce rappresentazioni di stati non-attuali. Ma essa non spiega — e per ragioni di principio non può spiegare — *che cosa significhi* per un soggetto esperire l'assenza come assenza. C'è qualcosa nel vuoto intenzionale husserliano — nel momento in cui si cerca il bicchiere sul tavolo dove dovrebbe essere e non c'è — che è

un'esperienza qualitativa specifica: il vuoto ha una forma, una direzione, una presenza paradossale che è categorialmente diversa dalla semplice assenza di percezione. Questo qualia dell'assenza — la fenomenologia specifica dell'intenzione vuota — è il residuo irriducibile che nessuna descrizione funzionale cattura completamente.

Questo residuo non è un'obiezione alla proposta neuroscientifica: è la sua frontiera naturale, il confine che separa il dominio della spiegazione meccanicistica dal dominio dell'esperienza. La negazione, come il linguaggio, come la coscienza, come il tempo vissuto, ha un versante che è accessibile alla scienza e un versante che è accessibile soltanto alla filosofia e alla fenomenologia — non perché la scienza sia inadeguata, ma perché i livelli di descrizione sono irriducibili tra loro e nessuno di essi esaurisce il fenomeno.

Ciò che questa analisi ha guadagnato rispetto al punto di partenza è la possibilità di tracciare con precisione la frontiera: si sa ora *cosa* la spiegazione neurale spiega (i meccanismi del reality monitoring, la struttura della simulazione controfattuale, la gerarchia dei tipi di negazione, il loro sviluppo ontogenetico e la loro dissoluzione patologica) e *cosa* non spiega (il carattere qualitativo dell'esperienza del vuoto, il senso soggettivo dell'assenza come assenza). Questa precisione è un guadagno epistemico reale: è meglio sapere con chiarezza dove si trova il limite di una teoria che procedere oltre quel limite con illusoria sicurezza.

Il paradosso della negazione — come il paradosso del linguaggio, come il problema difficile della coscienza — non è un problema che si risolve: è un problema che si *abita*, con gli strumenti progressivamente affinati che la scienza e la filosofia mettono a disposizione. Abitarlo con chiarezza e onestà intellettuale è forse la forma più alta di pensiero critico — che è, non casualmente, la forma di pensiero che la negazione rende possibile.

Negazione modale, mondi possibili e la topologia del necessariamente-assente

XXV. Dalla negazione semplice alla negazione modale: il salto dall'accidentale al necessario

Tutto ciò che si è detto finora sulla negazione proposizionale — la sua struttura husserliana di posizione-e-cancellazione, la sua implementazione come differenziale di precisione tra modello attuale e simulazione controfattuale, la sua stratificazione ontogenetica in tre fasi — si riferisce, implicitamente, a un tipo specifico di negazione: la negazione dell'accidentalmente non-vero. "Mario non è andato al cinema ieri" nega un evento che avrebbe potuto accadere e non è accaduto; "la mela sul tavolo non è rossa" nega una proprietà che l'oggetto potrebbe avere ma non ha. In entrambi i casi, il mondo controfattuale in cui la proposizione negata è vera è un mondo accessibile — un mondo compatibile

con le leggi fisiche e con la struttura causale dell'universo, differente dall'attuale soltanto per la specifica contingenza negata.

Esiste però una classe di negazioni di struttura radicalmente diversa: le negazioni del necessariamente-non-vero, le negazioni del logicamente o metafisicamente impossibile. "Un triangolo non può avere angoli la cui somma superi 180 gradi in geometria euclidea", "nessun numero pari maggiore di due è primo", "non esiste il maggiore numero primo" — queste non sono negazioni di contingenze: sono negazioni di impossibilità, affermazioni che nessun mondo accessibile, nessuna accumulazione futura di evidenze, nessuna simulazione ippocampale potrebbe mai soddisfare la proposizione negata. Nella terminologia della logica modale, si tratta di negazioni di proposizioni impossibili: $\Box \neg P$, ossia è necessariamente il caso che $\neg P$, il che equivale a $\neg \Diamond P$, ossia non è possibile che P .

La logica modale — nelle sue formulazioni classiche di C. I. Lewis, nella sua semantica dei mondi possibili di Saul Kripke e nella sua elaborazione per la filosofia della mente — offre il quadro formale naturale per distinguere questi due tipi di negazione e per portare questa distinzione all'interno del framework neuroscientifico. Ma la traiettoria da percorrere non è banale, e richiede di affrontare una tensione profonda che la letteratura neuroscientifica ha quasi completamente ignorato: la tensione tra la struttura semantica della modalità e la struttura computazionale del cervello probabilistico.

XXVI. La semantica dei mondi possibili e la sua anatomia neurale

Il framework semantico standard per la logica modale, dovuto a Kripke, è quello dei frame di Kripke già introdotti nella discussione sulla logica intuizionista, ma con una differenza cruciale. Nella semantica intuizionista, la relazione di accessibilità $\leq$ era un ordine parziale che rifletteva l'accumulazione dell'informazione: v è accessibile da w se v è epistemicamente più ricco di w . Nella semantica modale, la relazione di accessibilità R è più generale: $R(w, v)$ afferma semplicemente che il mondo v è accessibile dal mondo w nel senso rilevante per la modalità considerata — che può essere aletica (necessità e possibilità metafisica), deontica (obbligo e permesso), epistematica (sapere e ignorare), o temporale (sempre e talvolta).

Per la necessità aletica, la relazione di accessibilità è la compatibilità con le leggi naturali o logiche: v è accessibile da w se v e w condividono le stesse leggi. Per la necessità logica nella logica S5 — il sistema modale più forte tra quelli comunemente considerati — la relazione di accessibilità è la relazione universale: ogni mondo è accessibile da ogni altro, il che significa che le verità necessarie sono vere in tutti i mondi senza eccezioni.

La valutazione della necessità e della possibilità in questo framework è:

$$V(\Box P, w) = 1 \text{ sse } \forall v : R(w, v) \rightarrow V(P, v) = 1$$

$$V(\Diamond P, w) = 1 \text{ sse } \exists v : R(w, v) \wedge V(P, v) = 1$$

La negazione proposizionale ordinaria opera al livello di un singolo mondo: $\neg P$ è vera in w se P è falsa in w . La negazione modale forte — $\Box \neg P$ — opera attraverso tutti i mondi accessibili: è vera in w se P è falsa in ogni mondo accessibile da w . La differenza tra $\neg P$ e $\Box \neg P$ è la differenza tra "non accade che P nel mondo attuale" e "non può accadere che P in nessun mondo possibile".

Questa distinzione ha un correlato neurale preciso che la letteratura non ha ancora formulato esplicitamente, benché le evidenze empiriche puntino nella sua direzione. La negazione accidentale $\neg P$ — del tipo "Mario non è andato al cinema" — richiede la costruzione di un modello controfattuale ippocampale in cui Mario va al cinema: un mondo accessibile dal modello attuale, simulabile perché compatibile con le strutture causali che il modello generativo ha già appreso. La negazione necessaria $\Box \neg P$ — del tipo "un triangolo non può avere quattro lati" — non richiede questa simulazione, e in effetti non tollera di essere elaborata come simulazione: il sistema non può costruire un modello controfattuale in cui un triangolo ha quattro lati perché la struttura del concetto stesso blocca la simulazione. Tentare di costruire questa simulazione produce non un attrattore stabile nello spazio degli stati del modello generativo, ma un segnale di incoerenza strutturale — una forma di errore di previsione di secondo ordine che segnala non la discrepanza tra previsione e input, ma la discrepanza tra l'input concettuale richiesto e le strutture definitorie che il sistema ha consolidato.

L'analogia neurobiologica più precisa per questo secondo tipo di segnale è l'errore di previsione semantico documentato negli studi ERP con la componente N400: la risposta elettrofisiologica a violazioni semantiche ("ho bevuto il mio caffè con due zollette di cemento") è distinta dalla risposta a violazioni sintattiche (P600) e dalla risposta a violazioni pragmatiche o contestuali. Ma c'è una differenza importante tra la N400 prodotta da una violazione semantica contestuale e quella prodotta da una violazione concettuale necessaria. La prima — "il chirurgo ha operato con un bisturi/martello" — riflette la discrepanza tra aspettativa contestuale e parola ricevuta: è, strutturalmente, un errore di previsione di primo ordine, sensibile al contesto e alla frequenza. La seconda — "il triangolo aveva quattro lati" — riflette la violazione di una definizione costitutiva: è un errore di previsione di secondo ordine, insensibile al contesto e dipendente dalla struttura interna del concetto.

Questa distinzione — tra N400 contestuale e N400 definitoria — è stato documentata, in modo ancora embrionale, da studi di Hagoort e collaboratori su violazioni semantiche di diversa profondità concettuale, e da ricerche più recenti di Pykkänen sulla decomposizione logica delle frasi complesse tramite magnetoencefalografia. Ma la sua interpretazione modale — come differenza tra negazione accidentale e negazione necessaria — non è stata proposta. Essa suggerisce che il cervello operi implicitamente con una distinzione tra mondi

accessibili e mondi inaccessibili, e che questa distinzione sia implementata nella struttura del modello generativo attraverso due meccanismi distinti: la simulazione ippocampale per i mondi accessibili, e il segnale di incoerenza strutturale prefrontale per i mondi inaccessibili.

XXVII. Sistemi modali e vincoli sulle relazioni di accessibilità neurale

I principali sistemi della logica modale proposizionale sono caratterizzati da proprietà della relazione di accessibilità R che corrispondono ad assiomi supplementari aggiunti alla logica modale di base K . I sistemi principali sono:

K: nessuna proprietà di R richiesta; schema di validità: $\Box(P \rightarrow Q) \rightarrow (\Box P \rightarrow \Box Q)$.

T: R è riflessiva, ossia $R(w, w)$ per ogni w ; assioma corrispondente: $\Box P \rightarrow P$ (ciò che è necessario è vero).

S4: R è riflessiva e transitiva; assioma aggiuntivo: $\Box P \rightarrow \Box \Box P$ (ciò che è necessario, è necessariamente necessario).

S5: R è un'equivalenza (riflessiva, simmetrica, transitiva); assioma aggiuntivo: $\Diamond P \rightarrow \Box \Diamond P$ (ciò che è possibile, è necessariamente possibile).

B: R è riflessiva e simmetrica; assioma: $P \rightarrow \Box \Diamond P$ (ciò che è vero è necessariamente possibile).

Ciascuno di questi sistemi corrisponde a una famiglia di intuizioni diverse sulla struttura della modalità. S5 è il sistema standard per la necessità logica e matematica; S4 è adeguato per la necessità epistémica (la conoscenza è monotona: ciò che si sa rimane saputo); T è il sistema minimale per la modalità aletica.

La domanda che si può porre al cervello è: quale di questi sistemi descrive la struttura implicita della modalità nei processi cognitivi umani? Non si tratta di una domanda sulla logica formale che i parlanti conoscono esplicitamente — quasi nessuno ha familiarità con gli assiomi modali — ma di una domanda sulla struttura computazionale implicita dei processi di ragionamento modale.

Le evidenze psicologiche, benché non formulate in questi termini, suggeriscono che il ragionamento modale umano non sia omogeneo rispetto ai sistemi: gli esseri umani sono ragionatori S5 per la necessità logica e matematica (trattano le verità matematiche come necessarie in tutti i mondi possibili), ma ragionatori più deboli — probabilmente T o B — per la necessità fisica e causale (ammettono mondi in cui le leggi fisiche siano parzialmente diverse, come dimostrano le intuizioni sulla controfattualità causale). Questa eterogeneità sistemica ha una spiegazione naturale nel framework del modello generativo gerarchico: le

strutture concettuali più profonde e definitorie — quelle che corrispondono alle proprietà necessarie degli oggetti — sono rappresentate nei livelli più alti della gerarchia, dove i prior sono più rigidi e la plasticità sinaptica è più lenta; le strutture causali contingenti sono rappresentate nei livelli intermedi, dove l'aggiornamento bayesiano è più fluido.

La relazione di accessibilità del modello generativo neurale non è dunque unica, ma stratificata: diversi livelli della gerarchia corticale implementano relazioni di accessibilità con proprietà strutturali diverse, il che corrisponde al fatto che la modalità — la distinzione tra ciò che è necessario, ciò che è possibile e ciò che è contingente — non è una proprietà unificata del sistema cognitivo, ma una famiglia di proprietà che variano con il livello di astrazione concettuale.

Formalmente, si può esprimere questa stratificazione nel seguente modo. Sia $\mathcal{H} = \{h_1, h_2, \dots, h\}$ la gerarchia di livelli del modello generativo corticale, con h_1 il livello più basso (sensoriale) e h il livello più alto (concettuale-astratto). Per ogni livello h , si definisce una relazione di accessibilità R sullo spazio degli stati latenti al livello l . La proposizione $\Box P$ — necessità al livello l — è vera in uno stato w se P è vera in tutti gli stati v tali che $R(w, v)$. Le proprietà formali di R variano con l :

- Per l piccolo (livelli sensoriali-percettivi): R è densa e ha connettività alta — molti stati sono mutualmente accessibili, il che corrisponde alla plasticità percettiva e alla sensibilità contestuale.
- Per l grande (livelli concettuali-definitóri): R è sparsa, con cluster di accessibilità separati da barriere di potenziale elevate — pochi stati sono accessibili l'uno all'altro, il che corrisponde alla rigidità delle definizioni concettuali e all'inviolabilità delle proprietà necessarie.

La negazione necessaria $\Box \neg P$ al livello più alto corrisponde a P falsa in tutti gli stati del modello generativo in quel cluster concettuale: è la negazione delle proprietà definitorie, insensibile alla variazione contestuale e all'aggiornamento bayesiano locale.

XXVIII. La negazione deontica e il cervello normativo

Una dimensione della logica modale che ha una rilevanza particolare per la psicologia e la neuroscienza cognitiva è la modalità deontica — la logica dell'obbligo e della permissione. Le negazioni deontiche — "non si deve mentire", "non è permesso attraversare col rosso", "non è obbligatorio partecipare" — sono negazioni di norme, non di fatti, e il loro statuto ontologico è radicalmente diverso da quello delle negazioni aletiche. Un fatto negato non esiste nel mondo; una norma negata continua a esistere come norma — ciò che è proibito rimane una possibilità fattuale, semplicemente vietata.

Il sistema deontico standard SDL (Standard Deontic Logic) introduce gli operatori O (è obbligatorio che), P (è permesso che) e F (è proibito che), con le seguenti relazioni:

$$F P \equiv O \neg P$$

$$P P \equiv \neg O \neg P \equiv \neg F P$$

La negazione deontica $\neg O P$ — non è obbligatorio che P — è distinta sia dalla negazione fattuale $\neg P$ sia dalla proibizione $FP = O\neg P$. Questa distinzione triplice — tra il non-essere-vero ($\neg P$), il non-essere-obbligatorio ($\neg O P$), e l'essere-proibito ($O\neg P$) — è familiare a ogni parlante competente, ma la sua implementazione neurale è quasi completamente inesplorata.

Le evidenze disponibili suggeriscono un coinvolgimento distinto della corteccia prefrontale ventromediale e dell'insula nella elaborazione delle norme morali e sociali rispetto all'elaborazione dei fatti, con l'insula associata alla valutazione della violazione normativa e la vmPFC coinvolta nell'integrazione tra contenuto normativo e contesto situazionale. Ma la distinzione specifica tra la negazione di un obbligo e la proibizione corrispondente — la differenza neurale tra $\neg O P$ e $O\neg P$ — non è stata oggetto di studi diretti.

Questa distinzione ha però implicazioni cliniche rilevanti: la confusione tra "non è obbligatorio" e "è proibito" è una delle fonti più comuni di ragionamento morale distorto, ed è documentata in popolazioni con lesioni prefrontali che producono disinibizione comportamentale. La proposta è che questa confusione corrisponda a una perturbazione della struttura modale del modello generativo al livello normativo: il sistema perde la capacità di distinguere tra due configurazioni strutturalmente diverse della relazione di accessibilità deontica, collassando la distinzione tra $\neg O$ e $O\neg$ verso un'unica rappresentazione indifferenziata del "non-obbligatorio".

XXIX. Il problema della necessità de re e de dicto nel cervello

La distinzione modale più sottile — e neurologicamente più interessante — tra necessità de re e necessità de dicto risale ad Aristotele e attraversa tutta la filosofia medievale prima di ricevere un trattamento formale nel Novecento attraverso Quine, Kripke e la semantica dei mondi possibili. La distinzione riguarda la portata dell'operatore di modalità rispetto ai termini referenziali della proposizione.

La necessità de dicto — $\Box(P)$ — afferma che la proposizione P è necessariamente vera: è una modalità che opera sulla proposizione come un tutto. La necessità de re — per un oggetto x, la proprietà F è necessariamente vera di x — afferma che un determinato oggetto ha necessariamente una certa proprietà, indipendentemente da come lo si descrive. La proposizione "il numero 9 è necessariamente dispari" può essere interpretata de dicto (la proposizione "9

è dispari" è necessariamente vera — il che è corretto) o de re (l'oggetto che di fatto è 9 ha necessariamente la proprietà dell'essere dispari — il che è controverso se lo stesso oggetto viene considerato anche come "il numero di pianeti" che contingentemente potrebbe non essere dispari).

Per il cervello, questa distinzione ha una realizzazione concreta nel modo in cui le proprietà sono attribuite agli oggetti nel modello generativo. Un oggetto nel modello generativo è rappresentato come un bundle di proprietà con diversi gradi di necessità: alcune proprietà sono parte della rappresentazione definitiva dell'oggetto (necessità de re: l'acqua ha necessariamente la formula H_2O , nel senso che qualsiasi sostanza con una formula diversa non sarebbe acqua), altre sono proprietà contingenti acquisite nell'esperienza (l'acqua nel mio bicchiere è attualmente a 20 gradi).

Neuralmente, questa distinzione corrisponde alla differenza tra le rappresentazioni consolidate nei livelli alti del modello generativo — dove i prior sono forti e la revisione è lenta, corrispondente alle proprietà necessarie de re — e le rappresentazioni nei livelli più bassi, dove l'aggiornamento bayesiano è rapido e le proprietà contingenti vengono facilmente aggiornate dall'input sensoriale.

La negazione de re di una proprietà necessaria — negare che l'acqua sia H_2O , o che il triangolo abbia angoli — produce il segnale di incoerenza strutturale già discusso: non un errore di previsione di primo ordine (l'input non corrisponde alla previsione), ma una violazione della struttura interna del concetto. La negazione de re di una proprietà contingente — negare che l'acqua nel bicchiere sia a 20 gradi — è un ordinario errore di previsione di primo ordine, aggiornabile senza disturbare la struttura concettuale dell'oggetto.

Questa distinzione ha una previsione empirica specifica: la latenza e la distribuzione topografica della componente ERP associata alla negazione de re di proprietà necessarie dovrebbe differire sistematicamente da quella associata alla negazione de re di proprietà contingenti — con la prima più tardiva, più ampia in distribuzione frontale, e più resistente al contesto, e la seconda più precoce, più sensibile al contesto pragmatico, e più correlata con l'attivazione ippocampale.

Il paradosso del mentitore, l'auto-riferimento e i limiti computazionali della negazione

XXX. La proposizione che nega se stessa: il mentitore e la sua fenomenologia cognitiva

C'è una proposizione che ha tormentato la logica per oltre duemila anni — attribuita dall'antichità a Epimenide di Creta, formalizzata nella sua forma canonica da Alfred Tarski, e portata alla sua radicalità teorica da Bertrand Russell

e Kurt Gödel — e che pone alla negazione una sfida che nessuna delle teorie discusse finora può affrontare completamente: la proposizione auto-referenziale che nega se stessa. "Questa proposizione è falsa." Se è vera, allora è falsa; se è falsa, allora è vera.

Il paradosso del mentitore non è un gioco di parole: è il sintomo di un limite strutturale dei sistemi formali sufficientemente espressivi da poter parlare di se stessi. Tarski ha dimostrato che nessun sistema formale coerente può contenere il proprio predicato di verità — che la verità di un linguaggio può essere definita soltanto in un metalinguaggio superiore. Gödel ha dimostrato che nessun sistema formale sufficientemente ricco da contenere l'aritmetica di Peano può essere simultaneamente coerente e completo — che esistono proposizioni vere del sistema che il sistema non può dimostrare. Entrambi questi risultati sono connessi strutturalmente con la negazione auto-referenziale: il predicato di indimostrabilità gödeliano è una forma sofisticata di negazione che una proposizione applica a se stessa.

La domanda neurologicamente rilevante non è se il cervello umano sia un sistema formale nel senso di Gödel — non lo è, per ragioni che si esamineranno — ma se le strutture cognitive che implementano la negazione abbiano limiti analoghi a quelli dei sistemi formali, e se questi limiti siano osservabili come pattern di comportamento o come segnali neurofisiologici caratteristici.

La risposta, anticipata dall'analisi della paraconsistenza nella sezione XIX, è che il cervello non affronta il paradosso del mentitore con i meccanismi di negazione proposizionale ordinaria: lo affronta attraverso una serie di strategie computazionali che evitano la ricorsione infinita, stabilizzando il sistema in uno stato di oscillazione controllata o di sospensione epistemica che corrisponde a ciò che la logica formale chiama indecidibilità.

XXXI. Auto-riferimento e gerarchia predicativa: la soluzione biologica

La soluzione formale standard al paradosso del mentitore — la teoria ramificata dei tipi di Russell, poi semplificata nella teoria dei tipi semplici, e infine sistematizzata nella gerarchia tarskiana dei metalinguaggi — consiste nell'imporre una stratificazione gerarchica al linguaggio che impedisce l'auto-riferimento: le proposizioni che parlano della verità o falsità di proposizioni appartengono a un livello linguistico superiore rispetto alle proposizioni di cui parlano, e non possono parlare di se stesse.

Questa soluzione è formalmente soddisfacente ma fenomenologicamente insoddisfacente: gli esseri umani costruiscono e comprendono proposizioni auto-referenziali senza collassare in contraddizione, anche se non le risolvono. "Questa frase contiene cinque parole" è auto-referenziale e vera (o falsa, a seconda del conteggio). "La frase che stai leggendo ora è scritta in italiano" è auto-referenziale e vera. "Nessuna frase è vera" è auto-referenziale e paradossale.

Il cervello distingue questi casi — li tratta diversamente, produce reazioni cognitive diverse — benché non abbia una teoria esplicita della gerarchia dei tipi.

Come? La risposta biologica alla stratificazione formale è la separazione funzionale tra sistemi che operano a livelli diversi della gerarchia cognitiva, implementata attraverso la separazione anatomica e la modulazione temporale dei processi. La negazione proposizionale di primo ordine — la negazione di proposizioni su fatti del mondo — è elaborata in circuiti distinti dalla negazione di secondo ordine — la negazione di proposizioni su proposizioni, o di proposizioni sulle credenze del soggetto stesso. La prima coinvolge prevalentemente le aree perisilviane e le corteccie prefrontali laterali; la seconda, come si è argomentato nella sezione sulla Teoria della Mente e sulla metacognizione, coinvolge in misura molto maggiore le corteccie prefrontali mediali e la giunzione temporo-parietale.

Questa separazione anatomica non è perfetta — le connessioni tra i sistemi sono dense e bidirezionali — ma è sufficiente a impedire la ricorsione incontrollata nella maggior parte delle situazioni ordinarie. Il paradosso del mentitore è cognitivamente problematico proprio perché forza il sistema a elaborare un'auto-referenza che attiva contemporaneamente i circuiti di primo e di secondo ordine, producendo una situazione in cui ciascun circuito genera un segnale di errore che il circuito dell'altro livello interpreta come input da negare, con il risultato di un'oscillazione tra stati che non converge verso un attrattore stabile.

Questa oscillazione non è patologica: è la risposta adattiva di un sistema paraconsistente a una struttura proposizionale genuinamente indecidibile. Il cervello non collassa in risposta al paradosso del mentitore perché la sua architettura non richiede una risposta binaria definitiva — può mantenere la struttura in uno stato di precisione distribuita bassa, corrispondente all'entropia massima sulla dimensione vero/falso, senza propagare questa indeterminatezza al resto delle proprie rappresentazioni.

XXXII. Formalizzazione dell'auto-riferimento nel modello generativo gerarchico

Si può formalizzare questa proposta nel linguaggio della codifica predittiva gerarchica. Sia la gerarchia a L livelli con stati latenti $x^{(l)}$ e errori di previsione $\epsilon^{(l)}$ ad ogni livello. La negazione proposizionale di primo ordine al livello l è definita, come nella sezione VII, come:

$$N^{(l)}[P] \equiv "P \text{ produce } \epsilon^{(l)} \text{ massimo indipendentemente dall'input nei livelli } l' < l"$$

La negazione di secondo ordine — la negazione di una credenza del soggetto stesso — richiede la rappresentazione esplicita del valore di credenza come oggetto di un livello superiore. Formalmente, sia $B_w(P) \in [0,1]$ la credenza del

soggetto w nella proposizione P , rappresentata come stato latente al livello $l+1$. La negazione di secondo ordine $\neg B_w(P)$ — la negazione della credenza — è:

$$N^{(l+1)}[B_w(P)] \equiv "B_w(P) \text{ produce } \varepsilon^{(l+1)} \text{ massimo nel modello metacognitivo}"$$

La proposizione auto-referenziale del mentitore — "questa proposizione è falsa", indicata con L — genera una ricorsione che in termini di stati latenti si esprime come:

$$B_w(L) = 1 - B_w(L)$$

Questa equazione fissa ha la soluzione unica $B_w(L) = 0.5$, corrispondente all'entropia massima sulla credenza in L : il sistema non può assegnare a L né credenza alta né credenza bassa senza generare un errore di previsione al livello metacognitivo. La soluzione cognitiva è proprio questa: mantenere $B_w(L) \approx 0.5$, ossia trattare la proposizione come intrinsecamente indeterminata.

Ma questa soluzione non è stabile nello stesso senso in cui è stabile la credenza in una proposizione contingente incerta: per una proposizione contingente, $B_w(P) = 0.5$ riflette ignoranza epistemica che potrebbe essere risolta da nuove evidenze; per il mentitore, $B_w(L) = 0.5$ riflette impossibilità strutturale di determinazione che nessuna evidenza potrebbe risolvere. Nel framework della codifica predittiva, questa differenza si esprime attraverso la struttura della distribuzione di credenza: per la proposizione contingente incerta, la distribuzione è piatta perché la varianza è alta per mancanza di informazione; per il mentitore, la distribuzione è piatta perché l'equazione di punto fisso impedisce qualsiasi asimmetria — qualunque deviazione da 0.5 produce immediatamente un segnale di errore che la riconduce verso 0.5.

Il correlato neurofisiologico di questa differenza dovrebbe essere misurabile: la credenza in una proposizione contingente incerta dovrebbe produrre fluttuazioni del segnale neurale con alta varianza temporale (il sistema esplora diversi valori di credenza sulla base delle evidenze parziali); la credenza in una proposizione auto-referenzialmente paradossale dovrebbe produrre fluttuazioni con bassa varianza ma alta frequenza — l'oscillazione controllata tra stati quasi-opposti che è il correlato neurale dell'indeterminatezza strutturale.

XXXIII. Gödel, incompletezza e i limiti della negazione proposizionale nel cervello

Il risultato di incompletezza di Gödel del 1931 è, nella sua essenza, una dimostrazione che i sistemi formali sufficientemente espressivi non possono negare tutte le proposizioni false che appartengono al loro dominio: esistono proposizioni vere ma indimostrabili del sistema, proposizioni che il sistema non può né affermare né negare attraverso i propri assiomi e le proprie regole.

Questo risultato, tecnicamente relativo ai sistemi formali, ha implicazioni per la comprensione della negazione biologica che non sono state esplorate con la necessaria profondità.

La costruzione gödeliana procede attraverso la numerazione di Gödel: un'assegnazione sistematica di numeri naturali a tutti i simboli, formule e dimostrazioni del sistema formale, che permette di parlare del sistema nel linguaggio del sistema stesso. Questa numerazione è il meccanismo tecnico che rende possibile l'auto-riferimento: la proposizione gödeliana *G* afferma, nella sua codifica numerica, "il numero *n* non è il numero di Gödel di una dimostrazione di questa formula" — una forma di auto-referenza negativa che produce una proposizione vera ma indimostrabile.

Per il cervello, la domanda è se esista un analogo biologico della numerazione di Gödel — un meccanismo che permetta al sistema di rappresentare se stesso come oggetto di elaborazione — e se questo meccanismo produca limiti analoghi alla incompletezza gödeliana nella capacità di negazione del sistema.

La risposta è complessa e richiede di distinguere due domande diverse. La prima domanda — se il cervello abbia un analogo della numerazione di Gödel — ha risposta affermativa: il sistema di rappresentazione del sé, implementato primariamente nelle corteccie prefrontali mediali e nella corteccia cingolata posteriore, è il dispositivo biologico che permette al sistema di rappresentare se stesso come oggetto. Quando si pensa "io non credo che questa proposizione sia vera", si sta eseguendo un'operazione analoga alla auto-referenza gödeliana: il sistema rappresenta una sua propria credenza come oggetto di valutazione negativa.

La seconda domanda — se questa auto-referenza produca incompletezza nel senso gödeliano — ha risposta più sfumata. Il cervello non è un sistema formale nel senso tecnico di Gödel: non ha assiomi fissi, non ha regole di inferenza determinate, non produce dimostrazioni nel senso logico-formale. Ma ha qualcosa di funzionalmente analogo: prior consolidati che fungono da assiomi impliciti del modello generativo, e processi di inferenza bayesiana che fungono da regole di derivazione. In questo senso funzionale, ci sono proposizioni che il modello generativo del soggetto non può né affermare né negare: non perché siano logicamente indecidibili, ma perché il modello non ha sufficiente ricchezza strutturale per risolverle.

Questo fenomeno — la negazione biologicamente inattuabile — ha una controparte psicologica precisa nella nozione di blind spot cognitivo: le strutture del modello generativo che il soggetto non può mettere in questione perché qualsiasi tentativo di negazione presuppone quelle strutture. Un soggetto non può negare con successo le proprie assunzioni di base sulle categorie del reale — spazio, tempo, causalità, persistenza degli oggetti — non perché queste siano necessariamente vere, ma perché il meccanismo stesso di negazione le presuppone come condizioni di possibilità. È un'incompletezza funzionale, non

formale: i limiti della negazione biologica non sono isomorfici ai limiti gödeliani, ma ne sono l'analogo naturale in un sistema che opera per aggiornamento continuo piuttosto che per derivazione formale.

XXXIV. La semantica dei giochi e la negazione come sfida

C'è un approccio alla semantica logica che ha una connessione sorprendentemente naturale con l'architettura neurale del ragionamento: la semantica dei giochi di Jaakko Hintikka, in cui il significato di una proposizione è determinato da un gioco di verifica tra due giocatori — il Verificatore, che cerca di dimostrare la proposizione, e il Falsificatore, che cerca di confutarla. In questo framework, la verità di P significa che il Verificatore ha una strategia vincente nel gioco per P ; la falsità di P significa che il Falsificatore ha una strategia vincente.

La negazione in semantica dei giochi è elegantissimamente semplice: $\neg P$ è il gioco P con i ruoli di Verificatore e Falsificatore scambiati. Negare P significa prendere il lato del Falsificatore rispetto a P — adottare la prospettiva di chi cerca di confutare piuttosto di chi cerca di verificare. Questa inversione di prospettiva è, concettualmente, molto più vicina all'esperienza cognitiva della negazione di quanto lo sia la semplice inversione del valore di verità: quando si nega una proposizione, ci si posiziona come avversario nei confronti delle sue condizioni di verifica.

Nel cervello, questa struttura game-semantica ha un correlato naturale nell'architettura dei sistemi di controllo e monitoraggio cognitivo. La corteccia cingolata anteriore, già menzionata come sistema di rilevazione dei conflitti, funziona in un senso rilevante come arbitro del gioco di verifica: monitora la competizione tra il segnale bottom-up (le evidenze che il Falsificatore può invocare contro la previsione) e il segnale top-down (le previsioni del Verificatore che il modello generativo genera). L'errore di previsione è, in questi termini, la mossa vincente del Falsificatore: il segnale che la previsione del Verificatore è stata confutata dall'input.

La negazione proposizionale, in questa lettura, è il risultato di una partita in cui il Falsificatore ha vinto in modo così sistematico e robusto da indurre il sistema a aggiornare permanentemente il proprio modello: non un singolo errore di previsione, ma una vittoria strutturale del Falsificatore che porta il Verificatore a abbandonare la proposizione negata come previsione del proprio modello. Questo aggiornamento — la transizione da un modello che prevede P a un modello che prevede $\neg P$ — è la negazione come evento cognitivo, distinta dalla negazione come struttura proposizionale: è il momento in cui la negazione si incarna nel paesaggio energetico del sistema.

La semantica dei giochi offre anche un modo naturale di pensare la negazione parziale — la negazione di una componente specifica di una proposizione complessa. Nella semantica dei giochi, la struttura della partita per una proposizione congiuntiva $P \wedge Q$ è tale che il Falsificatore può scegliere quale

congiunto attaccare: questa asimmetria nella struttura del gioco corrisponde alla focalizzazione prosodica e pragmatica che determina la portata della negazione nelle frasi naturali. La negazione focalizzata "non ho mangiato la TORTA" (implicando che qualcosa ho mangiato) corrisponde a una partita in cui il Falsificatore ha scelto di attaccare soltanto il complemento oggetto della proposizione, concedendo la verità del resto: è una vittoria parziale, localizzata, che lascia intatta la struttura proposizionale globale modificando soltanto il sotto-gioco che corrisponde all'elemento focalizzato.

XXXV. Negazione gödeliana e creatività: il pensiero che supera il proprio modello

C'è un aspetto dell'incompletezza gödeliana che ha raramente ricevuto l'attenzione che merita nella letteratura sulla cognizione: il suo lato creativo. Il teorema di incompletezza dimostra che qualsiasi sistema formale sufficientemente ricco può essere esteso — aggiungendo come assioma la proposizione gödeliana G , che il sistema non può dimostrare — producendo un sistema più ricco in cui G è dimostrabile. Questo processo di estensione può essere iterato indefinitamente, producendo una gerarchia ascendente di sistemi formali ciascuno dei quali supera i limiti di incompletezza del precedente.

Nella cognizione umana, questo processo ha un analogo che si manifesta come capacità di mettere in questione le proprie assunzioni di base — di negare non le proposizioni che il proprio modello può negare dall'interno, ma le strutture del modello stesso che determinano cosa può essere affermato e negato. Questa capacità meta-cognitiva di secondo ordine — la capacità di negare il proprio sistema di negazione — è ciò che rende possibile la revisione radicale delle credenze, la conversione intellettuale, il cambiamento di paradigma nel senso kuhniano.

Nel framework del modello generativo gerarchico, questa capacità corrisponde alla plasticità dei livelli alti della gerarchia — i livelli che implementano i prior più fondamentali e le strutture concettuali più consolidate. In condizioni normali, questi livelli hanno bassa plasticità: i loro prior sono robusti e resistenti all'aggiornamento, il che li rende affidabili come base del ragionamento. Ma in condizioni di evidenza sufficientemente forte e consistente — o in stati cognitivi particolari come quelli prodotti dalla meditazione intensa, dall'esperienza psichedelica, o dall'elaborazione di traumi profondi — anche i livelli alti diventano plastici, e la negazione può operare su strutture che normalmente non sono soggette a revisione.

Questa plasticità dei livelli alti è la base neurologica del cambiamento conceptual profondo: la capacità di negare le strutture che si credevano necessarie, di scoprire che ciò che sembrava un vincolo definitivo era in realtà un prior forte ma aggiornabile, di estendere il proprio sistema cognitivo incorporando proposizioni che il sistema precedente non poteva né affermare né negare. È, in

un senso preciso, la realizzazione biologica del processo gödeliano di estensione del sistema: la negazione che supera i propri limiti estendendo il sistema in cui opera.

Il costo di questa espansione non è trascurabile: la revisione dei prior fondamentali produce tipicamente una fase di destabilizzazione cognitiva intensa — la dissonanza cognitiva acuta, l'ansia epistemica, la perdita temporanea della coerenza del modello — che corrisponde al periodo in cui il sistema opera con un modello generativo parzialmente destrutturato, prima che la nuova struttura si consolidi. Questa fase di destabilizzazione è il correlato cognitivo della transizione di fase nel paesaggio energetico del sistema: il momento in cui il vecchio attrattore perde profondità e il sistema oscilla in una regione di bassa barriera di potenziale prima di convergere verso il nuovo attrattore.

La negazione gödeliana — la negazione delle proprie assunzioni di base — è dunque la forma più costosa e più trasformativa di negazione biologica. È anche, non casualmente, la forma di negazione che caratterizza il pensiero filosofico nella sua modalità più radicale: la messa in questione non delle proposizioni che si credono false, ma delle strutture attraverso cui si distingue il vero dal falso. È la negazione che nega il proprio fondamento — e che, proprio per questo, apre il sistema verso un orizzonte di comprensione più ricco e più articolato.

La pragmatica della negazione marcata e l'economia informazionale del cervello

XXXVI. La marcatezza della negazione: da Greenberg a Grice

Esiste un fatto semplice e perturbante sulla negazione nel linguaggio naturale che le teorie logiche ignorano sistematicamente e che le teorie neurolinguistiche hanno trattato soltanto come curiosità statistica: la negazione è marcata. In tutte le lingue naturali studiate — e lo studio interlinguistico sistematico copre ormai diverse migliaia di sistemi linguistici — la forma affermativa di una proposizione è non marcata, la forma negativa è marcata. La marcatezza si manifesta su almeno quattro dimensioni distinte che Greenberg aveva individuato come universali implicazionali: la forma negativa è più lunga dell'affermativa (contiene più morfemi o sillabe); è acquisita più tardi nel corso dello sviluppo; è più rara nel discorso spontaneo; è processata più lentamente e con maggiore costo computazionale.

Questi quattro fatti, presi insieme, non sono una coincidenza: sono la traccia linguistica di una proprietà computazionale fondamentale del sistema cognitivo. La negazione è marcata perché è un'operazione di secondo ordine su una struttura già costruita — come si è argomentato nel dettaglio seguendo la struttura husserliana — e questa secondarietà si riflette in ogni aspetto dell'interfaccia tra il sistema cognitivo e il sistema linguistico.

Ma la marcatezza della negazione ha anche una dimensione puramente informazionale che è distinta dalla sua complessità computazionale, e che offre una prospettiva complementare sul problema. In un discorso ordinario, un'affermazione è attesa: riempie uno slot informazionale che l'interlocutore si aspettava di riempire con un valore determinato. Una negazione è inattesa: non riempie uno slot informazionale, ma chiude uno slot lasciandolo vuoto — e questa chiusura negativa è informazionalmente meno ricca dell'affermazione, perché non specifica quale tra le infinite alternative al valore negato sia quella vera.

Questa asimmetria informazionale è stata formalizzata da Paul Grice attraverso la massima di quantità: un parlante deve fornire tanta informazione quanta è necessaria per lo scopo conversazionale, né di più né di meno. Una negazione nuda — "non è un cane" — viola tendenzialmente questa massima, perché esclude un'ipotesi senza indicare quale tra le alternative sia corretta. La negazione è pragmaticamente informativa soltanto quando il contesto conversazionale ha già ristretto il campo delle alternative a un insieme limitato, tale che l'esclusione di una sia sufficiente per determinare la vera risposta. "Stai pensando a un animale?" "No." — la negazione è massimamente informativa perché il campo delle alternative è già ridotto a due (animale / non-animale).

XXXVII. La teoria della rilevanza e il costo computazionale della negazione

La teoria della rilevanza di Sperber e Wilson offre un trattamento più preciso della pragmatica della negazione attraverso il concetto di effetti cognitivi e sforzo di elaborazione. Secondo questa teoria, la comunicazione è un'impresa in cui il parlante cerca di massimizzare il rapporto tra effetti cognitivi prodotti nell'ascoltatore e sforzo di elaborazione richiesto: un enunciato è rilevante nella misura in cui produce molti effetti cognitivi con poco sforzo.

La negazione è, in questo quadro, tendenzialmente meno rilevante dell'affermazione corrispondente perché il suo effetto cognitivo principale — l'eliminazione di un'ipotesi — richiede che l'ipotesi sia già presente nella mente dell'ascoltatore come possibilità saliente. Se l'ipotesi non è saliente, la negazione non elimina nulla di cognitivamente rilevante, e il suo contributo agli effetti cognitivi è trascurabile a fronte di uno sforzo di elaborazione invariato. Questo spiega perché la negazione sia tipicamente usata come risposta a un'aspettativa attribuita all'interlocutore — esplicita o implicita — piuttosto che come forma primaria di descrizione del mondo: "il gatto non è sul tappeto" è naturale come risposta a chi si aspetta di trovarlo lì, innaturale come primo enunciato di una descrizione della stanza.

Nel framework neuroscientifico, questa analisi pragmatica corrisponde a un'analisi della struttura degli errori di previsione comunicativi. La negazione è un segnale di errore di previsione interpersonale: comunica all'interlocutore che

la sua previsione — esplicitata nella domanda o implicita nel contesto — non è corretta. Perché il segnale sia efficace, la previsione deve essere attiva nel sistema cognitivo dell'ascoltatore al momento della ricezione: un errore di previsione rispetto a una previsione che non si aveva non trasmette informazione, esattamente come un segnale di errore di previsione nel sistema percettivo di un organismo non porta aggiornamento al modello se il modello non aveva formulato una previsione per quel dominio.

Questa connessione tra la pragmatica della negazione e la struttura dell'errore di previsione interpersonale ha una previsione neuroimaging specifica: la risposta neurale alla comprensione di una frase negativa dovrebbe essere modulata dalla forza della previsione dell'ascoltatore rispetto all'ipotesi negata — con risposte più ampie e più precise per negazioni di ipotesi altamente salienti e aspettative forti, e risposte più deboli e più diffuse per negazioni di ipotesi di bassa salienza. Questa modulazione dovrebbe riflettersi nell'ampiezza del segnale di errore di previsione semantico (la N400) e nella latenza dell'aggiornamento del modello.

XXXVIII. L'implicatura scalare della negazione e la logica del "non tutto"

Tra le implicature pragmatiche più studiate nella letteratura linguistica, le implicature scalari occupano un posto particolare per la loro connessione con la struttura logica della negazione. Un'implicatura scalare emerge quando un parlante usa un termine di forza semantica minore all'interno di una scala ordinata di termini di forza crescente, implicando così che i termini di forza maggiore non siano applicabili. "Ho bevuto alcuni dei vini in cantina" implica — per implicatura scalare, non per entailment logico — che non ho bevuto tutti i vini: se avessi bevuto tutti, avrei usato il termine di scala superiore.

La relazione tra implicatura scalare e negazione è bidirezionale. Da un lato, l'implicatura scalare è una forma di negazione implicita: affermare "alcuni" implica la negazione di "tutti". Dall'altro, le negazioni esplicite interagiscono con le scale semantiche in modi complessi: "non tutti i filosofi sono platonici" implica — di nuovo per implicatura scalare — che almeno alcuni lo sono, producendo la lettura "alcuni ma non tutti". Questa lettura non è una conseguenza logica della negazione universale $\neg \forall x(Fx)$: logicamente, $\neg \forall x(Fx)$ è compatibile con $\forall x(\neg Fx)$. Ma pragmaticamente, la negazione dell'universale implica la verità del particolare.

Nel cervello, questa interazione tra negazione logica e implicatura scalare produce una situazione computazionalmente interessante: il sistema deve elaborare contemporaneamente il contenuto proposizionale della frase (la negazione logica dell'universale) e il suo contenuto pragmatico (l'implicatura che alcune istanze positive esistono), mantenendo la distinzione tra le due componenti per poterle usare separatamente nei ragionamenti successivi. Studi di ERP di Nieuwland e van Berkum, e successivamente di Politzer-Ahles e

Fiorentino, hanno mostrato che questa elaborazione parallela produce pattern neurofisiologici distinti per la componente semantica e quella pragmatica, con la N400 sensibile alla plausibilità semantica e una componente tardiva (approssimativamente 500–700 millisecondi) sensibile alla coerenza pragmatica dell'implicatura nel contesto.

Il risultato più sorprendente di questa linea di ricerca è che il cervello può elaborare l'implicatura scalare anche quando essa è in contraddizione con il contenuto proposizionale esplicito — e può mantenere questa contraddizione in sospeso per un periodo sorprendentemente lungo prima di risolverla. Ciò è possibile perché l'implicatura non è parte del contenuto proposizionale della frase: è un'inferenza generata dal sistema pragmatico sulla base dei principi conversazionali, e questo sistema opera con una relativa indipendenza dal sistema di elaborazione semantica. Nel framework della codifica predittiva, questi due sistemi corrispondono a livelli diversi della gerarchia predittiva — il sistema semantico opera ai livelli intermedi, il sistema pragmatico-conversazionale opera ai livelli alti — e la loro integrazione avviene con una latenza che riflette il costo computazionale dell'aggiornamento inter-livello.

XXXIX. La negazione metalinguistica rivisitata: tra uso e menzione

Nella sezione II del saggio si è introdotta la distinzione di Laurence Horn tra negazione proposizionale e negazione metalinguistica — la negazione che non nega il contenuto di un atto linguistico ma la sua forma. Si può ora tornare su questa distinzione con strumenti più ricchi, e mostrare che essa non è semplicemente una distinzione pragmatica ma una distinzione computazionale profonda che coinvolge livelli diversi della gerarchia cognitiva.

La negazione proposizionale — "Mario non è arrivato" — opera sul contenuto proposizionale: nega il fatto descritto dalla proposizione. La negazione metalinguistica — "non è partito, è emigrato" — opera sulla scelta lessicale: non nega il fatto che il soggetto abbia lasciato il paese, ma protesta contro la descrizione di quest'azione con il termine "partire" anziché "emigrare", implicando che il livello di specificità o la connotazione del termine sia inadeguato al contesto. La differenza è cruciale: la negazione proposizionale è falsa se il fatto descritto esiste; la negazione metalinguistica non è falsa se il fatto esiste — è inappropriata rispetto alla descrizione.

Questa asimmetria valore-di-verità vs. appropriatezza pragmatica corrisponde, nella terminologia degli atti linguistici di Austin e Searle, alla distinzione tra fallimento proposizionale (la proposizione è falsa perché il fatto non esiste) e fallimento illocutivo (l'atto linguistico è inappropriato perché le sue condizioni di felicità non sono soddisfatte). Nel cervello, questa distinzione richiede un sistema che elabori non soltanto il contenuto proposizionale delle frasi ma le loro

condizioni di uso appropriato — un sistema che abbia accesso non solo alla semantica ma alla pragmatica istituzionale del linguaggio.

La proposta è che la negazione metalinguistica attivi specificamente le regioni associate all'elaborazione delle norme sociali e dei registri linguistici — le cortecce prefrontali mediali e le regioni della giunzione temporo-parietale associate alla Teoria della Mente — più di quanto le attivi la negazione proposizionale ordinaria. Questa differenza di attivazione sarebbe il correlato neurale della distinzione tra operare sulla proposizione (dominio semantico-logico, cortecce perisilviane e prefrontali laterali) e operare sulla relazione tra parlante, enunciato e contesto (dominio pragmatico-sociale, reti frontali mediali e temporali).

XL. Il principio di rilevanza comunicativa della negazione e la sua evoluzione

Si può ora formulare un principio generale che sintetizza le diverse dimensioni della pragmatica della negazione, e che ha conseguenze sia per la teoria della comunicazione sia per la teoria dell'evoluzione del linguaggio:

Il principio di rilevanza comunicativa della negazione: una negazione è comunicativamente efficace nella misura in cui l'ipotesi negata è saliente e attiva nel modello cognitivo dell'interlocutore al momento della ricezione, e nella misura in cui la negazione riduce l'incertezza dell'interlocutore rispetto a un'alternativa determinata.

Questo principio ha una conseguenza evolutiva importante. La negazione proposizionale — la forma più costosa computazionalmente — è anche la più dipendente dalla salienza delle ipotesi nel modello dell'interlocutore, e quindi la più dipendente da una Teoria della Mente sufficientemente sofisticata da modellare le credenze e le aspettative altrui. La negazione è comunicativamente efficace soltanto tra agenti capaci di rappresentare le credenze degli altri — il che spiega perché la negazione proposizionale piena sia una prerogativa delle specie con Teoria della Mente di secondo ordine, come si è argomentato nella sezione XV.

Questa connessione tra pragmatica della negazione e Teoria della Mente suggerisce che la funzione comunicativa originaria della negazione — la funzione che ha spinto la selezione verso sistemi cognitivi capaci di negazione proposizionale — non fosse la descrizione del mondo ma la correzione delle credenze altrui: la negazione come atto sociale di aggiornamento del modello cognitivo dell'interlocutore. In questa prospettiva, la negazione proposizionale è una tecnologia cognitiva selezionata per l'efficienza della comunicazione sociale — per la capacità di correggere rapidamente le false credenze degli interlocutori, riducendo la dispersione di risorse comportamentali verso azioni basate su premesse false.

Negazione, logica temporale e il cervello che anticipa l'assente

XLI. La logica temporale e la struttura delle negazioni tensate

La sesta famiglia di logiche non-classiche rilevante per la teoria della negazione nel cervello — accanto all'intuizionistica, alla lineare, alla modale, alla paraconsistente e alla dei topos — è la logica temporale, sviluppata da Arthur Prior negli anni Cinquanta e formalizzata nelle sue versioni computazionalmente trattabili da Edmund Clarke, Amir Pnueli e altri nell'ambito della verifica automatica dei sistemi. La logica temporale introduce operatori che permettono di quantificare sulle posizioni temporali: G (globally, sempre in futuro), F (finally, prima o poi in futuro), X (next, nella posizione temporale successiva), e i loro duali passati H, P, Y.

La negazione in logica temporale interagisce con questi operatori in modi che hanno conseguenze dirette per la comprensione delle frasi tensate nel linguaggio naturale. La negazione di una proposizione temporalmente quantificata non è equivalente alla proposizione che nega il quantificatore temporale: $\neg G P$ (non sempre P) non equivale a $G \neg P$ (sempre non-P); la prima è l'esistenza di un momento in cui P è falsa, la seconda è l'universalità della falsità di P. Questa distinzione — tra la negazione di una quantificazione temporale e la quantificazione di una negazione — è sistematicamente produttiva di ambiguità nelle frasi del linguaggio naturale, e la sua risoluzione richiede di determinare la portata relativa della negazione e degli operatori temporali.

"Non ho sempre detto la verità" può significare $\neg G P$ (ci sono stati momenti in cui non ho detto la verità — la lettura più naturale) o $G \neg P$ (non ho mai detto la verità — la lettura paradossalmente totale). "Non dirò mai bugie" può significare $G \neg P$ (in ogni futuro momento non mentirò) o $\neg F P$ (non esisterà mai un futuro momento in cui mento — semanticamente equivalente a $G \neg P$ in logica temporale lineare, ma diverso in logica temporale ramificata). La disambiguazione di queste strutture richiede l'integrazione di informazioni prosodiche, contestuali e pragmatiche che il cervello elabora in modo implicito e automatico, e i cui correlati neurali iniziano solo ora ad emergere nella letteratura.

XLII. Il cervello come sistema predittivo tensato: previsioni, ritenzioni e protenzioni

Il framework della codifica predittiva, come si è visto nel corso del saggio, privilegia implicitamente una struttura temporale specifica: il ciclo previsione-errore-aggiornamento opera nel presente, confrontando le previsioni correnti con l'input corrente. Ma questa struttura temporale è troppo semplificata per catturare la ricchezza della cognizione temporalmente estesa del

cervello umano — e questa semplificazione si fa più evidente precisamente nel caso della negazione temporale.

Husserl, nell'analisi della coscienza del tempo interno (*Vorlesungen zur Phänomenologie des inneren Zeitbewusstseins*, 1905–10), aveva distinto tre componenti della struttura temporale della percezione: la ritenzione, che mantiene attivo il passato appena trascorso come orizzonte temporale della percezione corrente; il *punctum praesens*, l'istante percettivo presente; e la protenzione, l'anticipazione del futuro immediato come struttura vuota che attende il proprio riempimento. Questa triade non è una metafora: è una struttura fenomenologica precisa che corrisponde, nell'analisi neuroscientifica più recente, alla struttura degli stati predittivi del sistema cognitivo estesi nel tempo.

La ritenzione husserliana corrisponde alla traccia mnestica a breve termine — il mantenimento dell'attivazione neurale dopo la cessazione dello stimolo, implementato attraverso il firing persistente nelle corteccie prefrontali durante il delay period dei compiti di memoria di lavoro. La protenzione corrisponde alle previsioni anticipatorie generate dal modello generativo — le aspettative sul futuro immediato che preparano il sistema a elaborare l'input atteso. Il *punctum praesens* è il momento in cui previsione e input si incontrano, producendo o conferma o errore di previsione.

In questo quadro, la negazione temporale ha tre versioni strutturalmente distinte che coinvolgono processi neurali diversi:

La negazione della ritenzione — la negazione di ciò che si è appena percepito ("non ha detto X, ha detto Y") — è una correzione della traccia mnestica a breve termine: richiede l'aggiornamento della rappresentazione mantenuta in memoria di lavoro, il che produce attivazione nelle corteccie prefrontali dorsolaterali con una latenza che riflette il costo dell'aggiornamento (tipicamente 300–500 millisecondi dopo la ricezione del segnale correttivo).

La negazione del *punctum praesens* — la negazione di ciò che si sta attualmente percependo ("non vedo nulla", detto mentre si è in presenza di stimoli visivi subliminali) — è la forma più paradossale di negazione, perché nega l'input corrente mentre il sistema lo elabora. Questa struttura è coinvolta nei fenomeni di soppressione attiva della percezione conscia e nelle illusioni negative (come le *afterimages negative*), e il suo correlato neurale include la soppressione top-down delle corteccie sensoriali primarie attraverso connessioni discendenti dalle corteccie prefrontali — la forma più diretta di negazione come inibizione che tuttavia non è riducibile alla semplice inibizione.

La negazione della protenzione — la negazione di ciò che si aspettava di percepire ("non è successo ciò che mi aspettavo") — è strutturalmente identica all'errore di previsione: è il segnale che la protenzione non è stata soddisfatta dall'input. In questo senso, ogni errore di previsione è una negazione protentica:

nega la previsione anticipatoria, comunicando al sistema che il futuro atteso non si è materializzato.

XLIII. La logica temporale ramificata e la rappresentazione neurale dei futuri possibili

La logica temporale lineare — in cui il tempo è una sequenza unica e ordinata di istanti — è inadeguata per modellare il ragionamento controfattuale e le deliberazioni future: se il futuro è unico e determinato, non ha senso parlare di futuri alternativi o di possibilità future. La logica temporale ramificata (Branching Time Temporal Logic, BTL), sviluppata da Prior e formalizzata da Belnap nella sua logica delle azioni e delle alternative, supera questa limitazione introducendo una struttura ad albero del tempo: il passato è unico (un'unica traiettoria che ha portato al presente), ma il futuro è ramificato (un albero di possibili traiettorie, di cui una sola si realizzerà).

In BTL, gli operatori temporali si sdoppiano: $A \ G \ P$ (per tutte le traiettorie future, sempre P) vs. $E \ G \ P$ (esiste una traiettoria futura in cui sempre P); $A \ F \ P$ (su tutte le traiettorie, prima o poi P) vs. $E \ F \ P$ (su qualche traiettoria, prima o poi P). La negazione di questi operatori ramificati ha una struttura molto più ricca di quella della logica temporale lineare:

$\neg A \ G \ P \equiv E \ F \ \neg P$ (non necessariamente sempre P : su qualche traiettoria, prima o poi $\neg P$)

$\neg E \ G \ P \equiv A \ F \ \neg P$ (su nessuna traiettoria sempre P : su tutte le traiettorie, prima o poi $\neg P$)

$\neg A \ F \ P \equiv E \ G \ \neg P$ (non necessariamente prima o poi P : su qualche traiettoria, sempre $\neg P$)

$\neg E \ F \ P \equiv A \ G \ \neg P$ (non su nessuna traiettoria prima o poi P : su tutte le traiettorie, sempre $\neg P$)

Queste equivalenze — le dualità di de Morgan per la logica temporale ramificata — non sono soltanto curiosità formali: sono la struttura logica del ragionamento su futuri possibili, e la loro violazione sistematica è documentata nella psicologia del ragionamento sotto forma di errori caratteristici nel pensiero probabilistico sul futuro.

Il cervello opera implicitamente con una struttura analoga alla BTL: la corteccia prefrontale, l'ippocampo e il sistema dopaminergico mesolimbico costituiscono un circuito di simulazione di futuri possibili che corrisponde funzionalmente all'albero temporale ramificato. Diversi studi di neuroimaging — tra cui in modo particolarmente rigoroso quelli di Buckner e colleghi sulla corteccia entorinale come sistema di codifica degli spazi di stati, e quelli di Doll e Frank sull'interazione tra ippocampo e striato nel model-based planning — mostrano che il cervello mantiene rappresentazioni di traiettorie future multiple e valuta la loro probabilità relativa attraverso segnali di previsione del valore.

La negazione di una traiettoria futura — "non farò X", "non accadrà Y" — nel framework BTL è la chiusura di una ramificazione dell'albero temporale: il sistema elimina dalla propria rappresentazione delle possibilità future una traiettoria che era precedentemente accessibile. Questo non è semplicemente l'aggiornamento di una credenza su un fatto presente: è la modifica della struttura dell'albero temporale del modello generativo, con effetti che si propagano a tutte le inferenze future che dipendevano dalla traiettoria eliminata.

XLIV. L'asimmetria temporale della negazione e l'entropia informazionale

C'è una proprietà della negazione temporale che non ha equivalente nella negazione atemporale: la sua asimmetria rispetto alla direzione del tempo. La negazione del passato e la negazione del futuro non sono operazioni simmetriche — non soltanto computazionalmente, come si è accennato nella sezione XXII del saggio originale, ma informazionalmente e ontologicamente.

La negazione del passato opera su eventi che hanno già un valore di verità definitivo: il passato è determinato, e negare un evento passato significa correggere la propria credenza su un fatto già avvenuto. Questa correzione è possibile, ma ha un costo mnemonico che cresce con l'età della credenza: le credenze più vecchie sono più consolidate, più integrate in reti di credenze coerenti, e più costose da aggiornare. Formalmente, il costo dell'aggiornamento di una credenza sul passato è proporzionale alla sua entropia di Fisher rispetto alle evidenze disponibili — una misura della stabilità dell'attrattore nel paesaggio energetico del sistema.

La negazione del futuro opera su eventi che non hanno ancora un valore di verità: il futuro è aperto, e negare un evento futuro significa chiudere una possibilità che era ancora accessibile. Questa chiusura è anche più radicale della correzione del passato: elimina un futuro possibile dalla propria rappresentazione, modificando la struttura dell'albero temporale e tutte le aspettative che dipendevano da quella ramificazione. Il costo di questa chiusura è proporzionale all'importanza della traiettoria futura eliminata nel sistema di pianificazione del soggetto — alla profondità dell'attrattore che quella traiettoria costituiva nello spazio degli stati futuri possibili.

L'asimmetria temporale fondamentale — passato determinato vs. futuro aperto — ha un correlato nell'architettura neurobiologica del sistema di memoria e pianificazione. Il sistema ippocampale, che è centrale sia per la memoria episodica del passato sia per la simulazione del futuro, opera diversamente per le due direzioni temporali: per il passato, le sue rappresentazioni sono vincolate dalla traccia mnemonica consolidata e dalla sua integrazione con la conoscenza semantica; per il futuro, le sue rappresentazioni sono più plastiche e più variabili, riflettendo la genuina apertura del futuro come spazio di possibilità non ancora chiuse.

La negazione che opera su una traccia mnestica consolidata — "non era così, ricordavo male" — produce un processo di riconsolidamento attivo in cui la memoria viene resa transitoriamente plastica prima di essere ri-consolidata in una forma aggiornata. Questo processo è energeticamente costoso e produce un'esperienza soggettiva caratteristica di instabilità cognitiva temporanea — la sensazione di "riorganizzare" una credenza radicata. La negazione che opera su una proiezione futura — "non andrà così, devo rivedere le mie aspettative" — è meno costosa quando le aspettative sono recenti e poco integrate, più costosa quando sono profondamente radicate nei piani e nelle strutture motivazionali del soggetto.

XLV. Il lutto come negazione del futuro atteso: una proposta teorica

Tra i fenomeni psicologici che coinvolgono la negazione temporale, quello che ha la maggiore rilevanza clinica e teorica è il lutto — non soltanto il lutto per la perdita di una persona cara, ma il lutto come struttura cognitivo-affettiva che emerge ogni volta che un futuro fortemente atteso viene eliminato dall'albero temporale del soggetto. Il lutto è, nella sua essenza cognitiva, la negazione di un insieme di traiettorie future che costituivano una parte centrale del modello generativo del soggetto.

Il processo del lutto — nelle sue diverse fasi documentate clinicamente, dalla negazione iniziale alla depressione, dalla rabbia all'accettazione, nella sequenza proposta da Kübler-Ross e successivamente rivisitata dalla letteratura più recente che ne ha relativizzato l'universalità — è il processo attraverso cui il sistema cognitivo integra la chiusura di un insieme di traiettorie future nel proprio modello generativo. Ogni fase corrisponde a un diverso meccanismo di gestione della negazione temporale:

La fase di negazione iniziale — "non è vero, non può essere accaduto" — è il rifiuto dell'aggiornamento: il sistema riceve un segnale di errore di previsione di grandissima ampiezza (la morte, la perdita, la diagnosi) e tenta di sopprimerlo attraverso la soppressione dell'input piuttosto che attraverso l'aggiornamento del modello. È la forma di negazione che nella patologia corrisponde alla scotomizzazione psicoanalitica: la negazione che opera non sulla proposizione ma sulla sua formazione.

La fase depressiva — il ritiro, la tristezza pervasiva, la perdita di interesse per le attività quotidiane — corrisponde al processo di smantellamento delle strutture anticipatorie del modello generativo: il sistema deve eliminare le protenzioni che dipendevano dalle traiettorie future chiuse, e questo smantellamento produce una riduzione generalizzata dell'attività anticipatoria che fenomenologicamente si manifesta come perdita di senso e di interesse. Dal punto di vista della codifica predittiva, questa fase corrisponde a una riduzione della profondità delle previsioni — il sistema non genera più previsioni a lungo termine nelle aree del

futuro chiuso — con la conseguenza di un abbassamento generale del tono dell'attività mentale orientata al futuro.

L'accettazione — la fase finale del processo del lutto nei modelli che la prevedono — è l'integrazione della negazione nel modello generativo aggiornato: il sistema ha riconfigurato il proprio albero temporale eliminando le traiettorie chiuse e aprendosi a nuove traiettorie alternative. Questa integrazione non è la cancellazione del passato ma la sua ri-contestualizzazione: le traiettorie chiuse rimangono nel modello come traiettorie storicamente possibili ma ora definitivamente non-attuali — il che corrisponde, nella struttura formale della BTL, al loro spostamento dal futuro aperto al passato determinato: diventano fatti su ciò che non è stato.

La proposta è che il processo del lutto sia, nella sua struttura profonda, l'elaborazione biologica più estesa e più costosa della negazione temporale: la riscrittura del proprio albero temporale futuro in risposta alla chiusura forzata di ramificazioni centrali. Questa proposta ha conseguenze terapeutiche: il trattamento efficace del lutto patologico dovrebbe mirare non alla soppressione delle emozioni associate alla perdita, ma alla facilitazione del processo di riconfigurazione dell'albero temporale — aiutando il soggetto a costruire nuove traiettorie future che sostituiscano quelle chiuse senza negare la realtà della chiusura. È, in termini del framework neuroscientifico, un processo di aggiornamento del modello generativo guidato da scaffolding terapeutico: il terapeuta come collaboratore nella ricostruzione dell'albero temporale del soggetto.

XLVI. Negazione, anticipazione e la struttura dell'identità narrativa

Il punto d'arrivo di questa analisi della negazione temporale è il concetto di identità narrativa — la struttura cognitiva attraverso cui il soggetto si costituisce come agente nel tempo, integrando il passato ricordato, il presente percepito e il futuro anticipato in una narrazione coerente di sé. L'identità narrativa, nel senso di Paul Ricœur e dei suoi sviluppi in psicologia cognitiva, non è soltanto una struttura descrittiva retrospettiva: è una struttura anticipatoria che proietta il soggetto verso futuri possibili compatibili con la propria storia e con i propri valori.

La negazione ha un ruolo centrale nella costituzione dell'identità narrativa attraverso quello che si può chiamare il principio di esclusione identitaria: la propria identità è definita non soltanto da ciò che si è, ma da ciò che si esclude di essere — dai futuri che si chiude, dalle possibilità che si nega, dagli atti che si dichiara incompatibili con il proprio sé. "Non sono il tipo di persona che farebbe questo" è una negazione proposizionale che opera sull'identità del soggetto: nega che certe azioni appartengano all'albero temporale del sé, chiudendo ramificazioni del futuro in nome di una coerenza narrativa con il passato.

Questa negazione identitaria è computazionalmente la più complessa di tutte quelle esaminate nel saggio: richiede la simulazione controfattuale ippocampale di un sé alternativo (il sé che commetterebbe l'atto escluso), la valutazione della sua incompatibilità con il modello narrativo del sé attuale, e la marcatura di quella traiettoria come esclusa dall'albero temporale del soggetto — tutto questo in simultanea con il mantenimento della coerenza del modello generativo globale. È la negazione che coinvolge simultaneamente la memoria episodica (chi sono stato), il present modeling (chi sono ora) e l'anticipazione futura (chi sarò): la sintesi di tutti i tipi di negazione temporale in un'unica operazione che definisce il soggetto in rapporto alle proprie possibilità non-attualizzate.

L'identità non è ciò che si afferma di essere: è la struttura delle negazioni attraverso cui si definisce il campo delle proprie possibilità autentiche. In questo senso, la negazione non è soltanto una capacità cognitiva tra le altre — è la condizione di possibilità della soggettività stessa, intesa come progetto che si costituisce escludendo ciò che non è compatibile con il proprio essere-nel-tempo. Il cervello che nega non è soltanto un sistema di elaborazione proposizionale: è il substrato biologico di un soggetto che si costituisce attraverso le proprie esclusioni — che esiste, in parte e irriducibilmente, come la totalità dei futuri che ha scelto di non essere.

La negazione intersoggettiva, il corpo e la metafora concettuale

XLVII. La costituzione intersoggettiva della negazione: il "no" come atto sociale primario

C'è una dimensione della negazione che il saggio ha finora trattato come struttura del soggetto singolo — come operazione che un sistema cognitivo esegue sui propri contenuti rappresentazionali — e che richiede invece di essere pensata nella sua radice fondamentalmente intersoggettiva. La negazione, prima di essere un'operazione logica su proposizioni, è un atto sociale: il bambino di diciotto mesi che scuote la testa e dice no non sta eseguendo un'operazione computazionale sul proprio modello del mondo, ma sta compiendo un gesto comunicativo rivolto a un altro, modificando la sua azione nel mondo e regolando la relazione tra sé e l'altro. La negazione nasce nell'incontro, non nella solitudine del pensiero.

Questa priorità genetica della negazione intersoggettiva rispetto a quella intrasoggettiva non è semplicemente un fatto dello sviluppo ontogenetico — un accidente della traiettoria di acquisizione del linguaggio — ma riflette qualcosa di più profondo sulla natura stessa della negazione come struttura cognitiva. Michael Tomasello, nella sua ricostruzione dell'origine dell'unicità cognitiva umana, ha mostrato che la capacità di partecipare ad attività congiunte e di condividere stati mentali con un altro — la *shared intentionality* — è la matrice

dalla quale emergono le capacità linguistiche e simboliche specificamente umane. La negazione proposizionale, in questa prospettiva, è un'abilità che presuppone la condivisione di un terreno comune di credenze e aspettative con l'interlocutore: si nega non nel vuoto, ma rispetto a uno spazio di rappresentazioni condivise che la negazione va a perturbarne selettivamente una.

Questa priorità dell'intersoggettività ha un correlato neurobiologico specifico nel sistema dei neuroni specchio e nelle reti di simulazione motoria, ma soprattutto nell'architettura del sistema di comunicazione pre-verbale nell'infante. L'analisi di Colwyn Trevarthen sulla intersoggettività primaria — la capacità del neonato di sincronizzarsi con i pattern ritmici e affettivi del caregiver nelle prime settimane di vita — e quella di Daniel Stern sul processo di sintonizzazione affettiva mostrano che il sistema nervoso infantile è configurato, fin dalla nascita, per la ricezione e la produzione di segnali che regolano il comportamento altrui. Il gesto di rifiuto — la testa che scuote, la mano che respinge, la vocalizzazione prosodicamente discendente del no — emerge da questa infrastruttura di comunicazione corporea come sua prima differenziazione semantica: la distinzione tra il segnale di avvicinamento (accettazione, accordo, proseguimento) e il segnale di allontanamento (rifiuto, disaccordo, interruzione).

Neuroscientificamente, questa radice corporea della negazione ha un substrato nell'architettura del sistema motorio e delle sue proiezioni sulla corteccia prefrontale. La distinzione approccio-evitamento — che nell'animale non umano è puramente comportamentale e nella sua implementazione neurale corrisponde all'attivazione differenziale del nucleo accumbens (approccio, ricompensa attesa) e dell'amigdala basolaterale (evitamento, minaccia rilevata) — nel cervello umano viene progressivamente de-somatizzata e simbolizzata: la negatività della testa che scuote viene trasferita al morphema negativo, e il gesto corporeo diventa segno linguistico senza perdere la sua radice nella dinamica di approccio-evitamento. Questa de-somatizzazione non è mai completa: la negazione proposizionale adulta mantiene tracce della sua origine corporea nell'embodiment delle rappresentazioni negative (la tensione muscolare associata al rifiuto, la prosodica discendente delle frasi negative in tutte le lingue studiate, la tendenza a movimenti di allontanamento durante l'elaborazione di contenuti negativi documentata da studi di compatibilità azione-parola).

XLVIII. La pragmatica del "no" congiunto: negazione condivisa e spazio mentale comune

C'è una forma di negazione che presuppone in modo particolarmente esplicito la struttura intersoggettiva: la negazione congiunta, l'atto di negare insieme a un altro qualcosa che entrambi hanno fino a quel momento creduto. "No, stavamo sbagliando" — questo plurale non è retorico: è la registrazione di un aggiornamento del modello condiviso tra due o più agenti cognitivi. La negazione congiunta è un atto di allineamento verso il basso — non si aggiorna un modello individuale, ma si modifica uno spazio mentale comune nel senso di

Fauconnier, uno spazio di rappresentazioni che entrambi i partecipanti al discorso trattano come condivise.

La teoria degli spazi mentali di Gilles Fauconnier — sviluppata poi in connessione con la teoria dell'integrazione concettuale (conceptual blending) di Fauconnier e Turner — offre un quadro in cui la negazione è la perturbazione selettiva di uno spazio mentale: non cancella l'intera struttura dello spazio, ma modifica la mappatura tra elementi dello spazio e il dominio cui si applica. "Il presidente non è a Parigi" modifica lo spazio mentale in cui si stavano tracciando inferenze sull'ubicazione del presidente, eliminando dalla mappatura l'associazione presidente-Parigi senza annullare l'intero spazio (il presidente continua ad esistere come elemento dello spazio; Parigi continua ad esistere come luogo nello spazio; soltanto la loro associazione viene sbarrata).

Questa struttura di perturbazione selettiva degli spazi mentali ha un correlato neurobiologico nell'architettura distribuita delle rappresentazioni semantiche: un oggetto è rappresentato come un bundle di proprietà distribuite attraverso diverse aree corticali (colore nelle aree visive ventrali, localizzazione nelle aree parietali, azione associata nelle aree premotorie, identità semantica nelle aree temporali), e la negazione di una singola proprietà non annulla il bundle ma elimina o sopprime selettivamente uno dei suoi componenti mantenendo intatti gli altri. La selettività della negazione — il fatto che negare "il presidente è a Parigi" non cancelli la rappresentazione del presidente né quella di Parigi — riflette questa architettura distribuita in cui la negazione può operare su sotto-reti specifiche del bundle rappresentazionale senza coinvolgere l'intero sistema.

Nella negazione congiunta, questa selettività deve essere coordinata tra i due agenti: entrambi devono identificare quale sotto-rete della rappresentazione condivisa viene modificata dalla negazione, e devono farlo in modo compatibile. Questo richiede non soltanto la comprensione proposizionale della frase negativa, ma la sua contestualizzazione nello spazio mentale condiviso — un'operazione che coinvolge il sistema di Teoria della Mente e la capacità di simulare lo spazio concettuale dell'interlocutore per identificare quale mappatura stia modificando. Il costo computazionale della negazione congiunta — documentato in studi di tempi di reazione su compiti di comprensione del discorso dialogico — è sistematicamente maggiore di quello della negazione in monologue, confermando che la coordinazione intersoggettiva aggiunge un livello di elaborazione rispetto alla negazione individuale.

XLIX. La negazione e il corpo: metafore concettuali e il cervello incarnato

La teoria della metafora concettuale di George Lakoff e Mark Johnson ha mostrato, a partire da *Metaphors We Live By* (1980) e nel successivo lavoro di Lakoff in *Philosophy in the Flesh* (1999), che le strutture concettuali astratte — incluse quelle logiche e matematiche — non sono indipendenti dal corpo, ma

sono strutturate attraverso schemi di immagini (image schemas) che derivano dall'esperienza corporea e sensomotoria. La domanda naturale che questo approccio pone al problema della negazione è: esiste uno schema di immagini corporeo che struttura la negazione concettuale, e se sì quale?

La risposta che emerge dall'analisi crosslinguistica e dalla psicologia cognitiva è che la negazione è strutturata da almeno tre schemi di immagini fondamentali, ciascuno dei quali corrisponde a una delle sue manifestazioni corporee primarie. Il primo è lo schema di BLOCCO (blockage): qualcosa che avrebbe dovuto procedere non procede, perché un ostacolo si interpone tra l'agente e la meta. Le espressioni linguistiche di negazione come "non ce la fa", "non passa", "non arriva" sfruttano sistematicamente questa struttura di impedimento al movimento. Il secondo è lo schema di ALLONTANAMENTO o ESPULSIONE: un oggetto viene rimosso dallo spazio di pertinenza dell'agente. Le negazioni come "non è qui", "non c'entra", "non appartiene" sfruttano lo schema dell'allontanamento. Il terzo è lo schema di ASSENZA DI CONTENUTO: un contenitore che dovrebbe avere qualcosa risulta vuoto. Le negazioni come "non ho nulla", "non rimane niente", "è vuoto" sfruttano la struttura del contenitore privo di contenuto.

Questi tre schemi non sono metafore ornamentali: sono strutture cognitive che organizzano la comprensione della negazione a un livello preriflessivo, precedente all'elaborazione proposizionale esplicita. La loro base neurobiologica è nelle cortecce sensomotorie che implementano la simulazione di movimenti bloccati, di oggetti allontanati, di contenitori svuotati — le stesse cortecce che, come si è visto, vengono attivate durante la comprensione di frasi negative nel paradigma dell'embodied cognition. La negatività frontale tardiva (LFN) associata alla negazione potrebbe parzialmente riflettere non soltanto il costo computazionale della costruzione della struttura proposizionale negativa, ma l'attivazione e la successiva soppressione degli schemi sensomotori sottostanti che la negazione innesca prima di essere elaborata a livello proposizionale.

L'implicazione teorica di questa analisi è che la negazione proposizionale adulta non sia affatto astratta e disincarnata nel modo in cui la logica formale la rappresenta: è un'operazione concettuale che si appoggia continuamente su strutture sensomotorie incarnate, e la sua "purezza logica" è il risultato di un processo di astrazione che mantiene la connessione con la sua base corporea pur non dipendendo direttamente da essa per ogni atto negativo specifico. La negazione è incarnata non nel senso che per produrla si debbano muovere muscoli, ma nel senso che la struttura concettuale che la organizza è modellata dalla struttura delle interazioni corporee con il mondo — e questo modellamento lascia tracce nel cervello che sono misurabili e che non scompaiono con l'acquisizione della competenza proposizionale astratta.

L. Variazione crosslinguistica e universali della negazione: cosa dice la tipologia linguistica

La tipologia linguistica offre un laboratorio naturale per testare ipotesi sulla struttura cognitiva della negazione: se la negazione ha basi cognitive universali radicate nell'architettura neurobiologica della specie, ci si aspetta che le sue proprietà fondamentali siano invarianti attraverso le lingue; se invece la negazione è una struttura puramente arbitraria e convenzionale, ci si aspetta variazione non sistematica tra lingue diverse. La realtà, come sempre in questo campo, è più interessante di entrambe le ipotesi estreme.

L'universale più robusto è quello già discusso della marcatezza: la forma negativa è marcata rispetto all'affermativa in tutte le lingue studiate, senza eccezioni documentate. Questo fatto — 7000 lingue, zero eccezioni — non può essere spiegato convenzionalmente: riflette la struttura computazionale della negazione come operazione di secondo ordine sulla rappresentazione affermativa che la precede.

Il secondo universale — documentato da Matthew Dryer nel database World Atlas of Language Structures — è la correlazione tra l'ordine delle parole e la posizione della negazione: nelle lingue con ordine Soggetto-Verbo-Oggetto, la negazione si trova tipicamente dopo il soggetto e prima del verbo; nelle lingue con ordine Soggetto-Oggetto-Verbo, la negazione tende a seguire l'oggetto. Questa correlazione non è assoluta, ma è statisticamente fortissima e suggerisce che la posizione della negazione nella struttura lineare della frase sia non arbitraria ma vincolata dalla struttura dell'elaborazione temporale del linguaggio: la negazione deve arrivare all'ascoltatore nel momento in cui la rappresentazione proposizionale è sufficientemente costruita da poter essere modificata, e questo momento dipende dall'ordine in cui i costituenti vengono presentati.

La principale fonte di variazione crosslinguistica riguarda la struttura morfologica della negazione: alcune lingue realizzano la negazione attraverso un affisso sul verbo (morfologia flessiva), altre attraverso un avverbio separato (negazione analitica), altre ancora attraverso particelle clitiche. Johanson e Wichmann hanno mostrato che questa variazione non è casuale ma è correlata con la struttura prosodica e con la complessità morfologica complessiva della lingua. Ma ancora più interessante è la variazione nella struttura della negazione doppia: in alcune lingue (italiano, francese, russo), la negazione richiede la concordanza negativa — più elementi della frase portano la marca negativa, producendo strutture come "non ho visto nessuno" dove sia "non" sia "nessuno" sono portatori della negazione; in altre lingue (inglese standard), la doppia negazione è logicamente positiva ("I didn't see nobody" è non-standard e implicita l'affermazione "ho visto qualcuno").

La concordanza negativa è particolarmente interessante dal punto di vista cognitivo perché mostra che la negazione può diffondersi attraverso la struttura sintattica della frase come una caratteristica di accordo, esattamente come il genere grammaticale o il caso. Questo fenomeno suggerisce che la negazione non sia soltanto un operatore proposizionale puntuale, ma una proprietà strutturale che può essere codificata in modo distribuito attraverso la frase — il

che corrisponde, nel cervello, a una modulazione distribuita del segnale di elaborazione che non è localizzata in un singolo punto del processing ma si propaga attraverso la rete di elaborazione sintattica. La concordanza negativa sarebbe così il riflesso linguistico di una negazione neurale distribuita: non un singolo segnale inibitorio in un punto della catena di elaborazione, ma una modulazione globale del tono del processing che investe l'intera struttura sintattica.

Il principio di energia libera, la negazione come minimizzazione della sorpresa e una sintesi formale

LI. Il principio di energia libera e la negazione come struttura fondamentale dell'inferenza attiva

Il framework della codifica predittiva, che ha accompagnato l'intera analisi neuroscientifica del saggio, trova la sua formulazione più generale e formalmente più rigorosa nel principio di energia libera di Karl Friston — una proposta che ambisce non soltanto a descrivere i meccanismi dell'elaborazione sensoriale e motoria del cervello, ma a fornire un principio unificato per la comprensione di qualsiasi sistema autopoietico, ossia qualsiasi sistema che mantenga la propria integrità strutturale nel tempo attraverso l'interazione con l'ambiente. Portare il principio di energia libera alla teoria della negazione non è un'operazione meramente formale: rivela che la negazione, nella sua struttura più profonda, non è una capacità speciale del cervello umano aggiunta all'elaborazione di base, ma è implicata nella struttura stessa di qualsiasi sistema che si mantenga come sistema.

Il principio di energia libera afferma che qualsiasi sistema biologico autopoietico minimizza la sua sorpresa — tecnicamente, minimizza la bounded surprisal o variational free energy, che è un limite superiore alla sorpresa epistemica:

$$\begin{aligned} F &= E_q[\log q(\psi) - \log p(\psi, o)] \\ &= D_{\text{KL}}[q(\psi) \parallel p(\psi \mid o)] - \log p(o) \end{aligned}$$

dove ψ rappresenta gli stati interni del sistema (gli stati latenti del modello generativo), o rappresenta gli osservabili (l'input sensoriale), $q(\psi)$ è la distribuzione variazionale che il sistema mantiene come approssimazione della distribuzione posteriore $p(\psi \mid o)$, e D_{KL} è la divergenza di Kullback-Leibler. Minimizzare F equivale simultaneamente a minimizzare la divergenza tra il modello del sistema e la vera distribuzione posteriore, e a massimizzare la log-evidenza del modello — ossia a ottimizzare sia la precisione del modello sia la sua adeguatezza all'input.

In questo framework, la negazione emerge come struttura fondamentale della minimizzazione dell'energia libera attraverso un meccanismo che si può

descrivere nel modo seguente. Quando il modello del sistema prevede P con alta precisione e l'input segnala $\neg P$ — quando c'è una forte discrepanza tra la previsione del modello generativo e l'evidenza sensoriale — l'energia libera aumenta significativamente. La minimizzazione dell'energia libera può procedere attraverso due percorsi distinti, che Friston chiama *percepire* (updating the recognition model) e *agire* (acting to change the sensory input):

Nel percorso percettivo, il sistema aggiorna il proprio modello generativo in direzione $\neg P$: modifica i parametri del modello in modo da ridurre la discrepanza tra la previsione e l'input. Questo è ciò che si è chiamato in tutto il saggio la negazione come aggiornamento del modello — la forma basilare di negazione biologica.

Nel percorso attivo, il sistema agisce sull'ambiente in modo da modificare l'input sensoriale verso P : compie azioni che portano la realtà in accordo con la propria previsione piuttosto che aggiornare la previsione in accordo con la realtà. Questo percorso corrisponde a un rifiuto attivo della negazione — all'insistenza che il mondo sia come il modello prevede piuttosto che come l'evidenza suggerisce.

La dialettica tra questi due percorsi — accettare la negazione aggiornando il modello, o rifiutarla agendo sul mondo — è la struttura profonda di ogni deliberazione e di ogni apprendimento. L'essere umano maturo oscilla continuamente tra questi due modi di rispondere alla discrepanza tra aspettativa e realtà, e la qualità del suo adattamento dipende dalla capacità di scegliere il percorso appropriato al contesto: aggiornare il modello quando la realtà è imm modificabile, agire sul mondo quando il modello è corretto e la realtà è modificabile. Questa oscillazione non è semplicemente una strategia cognitiva: è la struttura fondamentale dell'esistenza come sistema autopoietico nel mondo.

LII. La negazione come riduzione dell'entropia del modello: una formalizzazione integrata

Il principio di energia libera permette di integrare in un unico framework formale tutti i tipi di negazione discussi nel saggio, attraverso il concetto di precisione (precision) già utilizzato nella sezione XII. La precisione di una previsione è l'inverso della sua varianza: $\Pi = \Sigma^{-1}$. Un modello generativo è caratterizzato, ad ogni livello l della gerarchia, da previsioni $\mu^{(l)}$ e da precisioni $\Pi^{(l)}$. L'energia libera al livello l è:

$$F^{(l)} = \frac{1}{2} \epsilon^{T(l)} \Pi^{(l)} \epsilon^{(l)} - \frac{1}{2} \log |\Pi^{(l)}| + \text{const}$$

dove $\epsilon^{(l)} = o^{(l)} - \mu^{(l)}$ è l'errore di previsione al livello l , ossia la discrepanza tra la previsione del modello e l'input (o la previsione del livello inferiore). L'errore di previsione pesato per la precisione — $\Pi^{(l)} \epsilon^{(l)}$ — è ciò che si propaga verso l'alto come segnale di aggiornamento: errori ampi in domini con alta precisione

producono aggiornamenti forti del modello; errori ampi in domini con bassa precisione producono aggiornamenti deboli.

In questo quadro, si possono ora caratterizzare formalmente i diversi tipi di negazione discussi nel saggio come operazioni distinte sul paesaggio di energia libera del sistema:

La **negazione percettiva primaria** — l'errore di previsione sensoriale — è un picco dell'energia libera al livello $l=1$ che si propaga verso l'alto come segnale di aggiornamento fino a trovare il livello della gerarchia sufficientemente plastico da assorbire la discrepanza. Formalmente: $\Delta F^{(1)} \gg 0$, con propagazione ascendente $\Pi^{(1)} \varepsilon^{(1)} \rightarrow x^{(2)}$.

La **negazione epistemica** — la rappresentazione esplicita che una credenza è errata — è un picco dell'energia libera al livello metacognitivo $l=k>1$, prodotto dall'integrazione degli errori di previsione attraverso i livelli inferiori. Richiede un modello esplicito delle proprie credenze come oggetti di valutazione: $F^{(k)} = D_KL[q(B_w) \parallel p(B_w \mid \text{evidenza})]$, dove B_w è la distribuzione di credenze del soggetto.

La **negazione logica piena** — l'operazione che nega una proposizione astratta indipendentemente dall'input — è una riduzione della precisione della rappresentazione di P al livello più alto della gerarchia, prodotta non dall'input sensoriale ma dalla struttura del modello generativo stesso: $\Pi^{(L)}(P) \rightarrow 0$, con aumento corrispondente della precisione di $\neg P$: $\Pi^{(L)}(\neg P) \rightarrow \Pi_max$.

La **negazione modale necessaria** — $\Box \neg P$ — è una configurazione del modello generativo in cui la riduzione della precisione di P è strutturalmente stabile: non c'è percorso di aggiornamento bayesiano che possa ripristinare $\Pi^{(L)}(P) > 0$, perché la struttura interna del modello non genera previsioni compatibili con P in nessun contesto.

La **negazione temporale futura** — la chiusura di una traiettoria nell'albero temporale ramificato — è una riduzione della precisione della previsione di P per tutti i timestep $t' > t_now$: $\forall t' > t, \Pi^{(L)}(P, t') \rightarrow 0$, con riallocazione della massa probabilistica verso le traiettorie alternative.

Questa formalizzazione unificata permette di esprimere la gerarchia dei tipi di negazione come una gerarchia di profondità dell'intervento sul modello generativo: la negazione percettiva è l'intervento più locale e reversibile (modifica un singolo livello, è facilmente annullabile); la negazione modale necessaria è l'intervento più globale e irreversibile (modifica la struttura topologica del modello, è resistente a qualsiasi aggiornamento bayesiano locale).

LIII. La negazione nei sistemi artificiali come specchio diagnostico della negazione biologica

L'analisi della negazione nei grandi modelli linguistici — i Large Language Models (LLM) basati sull'architettura transformer — offre un confronto diagnostico di notevole interesse teorico: non perché questi sistemi siano modelli del cervello umano, ma perché le loro difficoltà specifiche con la negazione rivelano, per contrasto, le proprietà della negazione biologica che non sono riproducibili attraverso l'addestramento statistico su testi.

I LLM mostrano una difficoltà sistematica e ben documentata con la negazione proposizionale in contesti che richiedono ragionamento multi-step. Questo fallimento non è casuale: riflette la struttura della loro architettura, che è fondamentalmente uno stimatore della distribuzione di probabilità della parola successiva dato il contesto — ossia un sistema statistico sulla distribuzione $p(\text{token}_{t+1} \mid \text{token}_1, \dots, \text{token}_t)$. In questa struttura, la negazione è appresa come pattern co-occorrenziale: il modello impara che certi pattern linguistici (frasi con "non", "mai", "nessuno") co-occorrono statisticamente con certe continuazioni. Ma la negazione proposizionale richiede qualcosa di diverso: richiede di costruire una rappresentazione del contenuto negato e di mantenerla attiva con una marcatura di non-attualità mentre si ragiona sulle sue implicazioni. Questa struttura dual-coding — contenuto presente e marcatura negativa presente simultaneamente — non emerge facilmente da un addestramento statistico sulla distribuzione dei token, perché richiede la costruzione di stati interni espliciti che i transformer non implementano in modo nativo.

Le difficoltà specifiche dei LLM con la negazione includono: la confusione tra negazione proposizionale e negazione metalinguistica (tendenza a interpretare "non ho detto che P" come "ho detto che non-P"); la difficoltà con negazioni annidate profonde ("non è falso che non sia impossibile"); la tendenza a dimenticare la negazione in catene di ragionamento lunghe, producendo conclusioni che trattano la premessa negata come affermativa; e la sensibilità al modo in cui la negazione viene formulata — con variazioni non semanticamente significative della formulazione che producono risposte radicalmente diverse.

Queste difficoltà riflettono l'assenza, nei LLM attuali, di tre componenti che nel cervello biologico implementano la negazione proposizionale robusta: la simulazione controfattuale ippocampale (che costruisce il mondo in cui P è vera per poi marcarlo come non-attuale), il reality monitoring prefrontale (che mantiene la distinzione tra reale e controfattuale nel corso del ragionamento), e la struttura gerarchica di errori di previsione pesati per la precisione (che permette di distinguere la robustezza delle negazioni sulla base della qualità delle prove che le supportano).

La negazione nei LLM è, in ultima analisi, una negazione senza vuoto intenzionale: i sistemi producono output negativi senza che ci sia un atto cognitivo che si dirige verso il contenuto negato come assente. La struttura husserliana della posizione-e-cancellazione — che richiede la costruzione esplicita della rappresentazione affermativa prima della sua modificazione negativa — non è implementata in sistemi che generano token sequenzialmente

senza costruire stati interni espliciti del contenuto proposizionale. Questo confronto non è un'osservazione sull'inadeguatezza dei LLM — sono sistemi potentissimi per altri scopi — ma un'indicazione di ciò che la negazione biologica richiede e che la negazione statistica non produce: la presenza attiva dell'assente, il vuoto con una forma.

LIV. Le filosofie orientali della negazione: il catuskoti, il prasanga e la logica del silenzio

La tradizione filosofica occidentale — greca, medievale, moderna, analitica — non è l'unico sistema di pensiero che abbia elaborato una teoria della negazione di qualità e profondità paragonabili a quelle discusse nel saggio. La filosofia indiana classica, in particolare il Madhyamaka buddhista di Nagarjuna (II secolo d.C.) e le sue elaborazioni da Chandrakirti a Tsongkhapa, ha sviluppato una teoria della negazione che non soltanto anticipa alcune delle strutture formali della logica paraconsistente, ma le arricchisce con una dimensione soteriologica e pratica che non ha equivalente nel pensiero analitico occidentale.

Il catuskoti — letteralmente "le quattro corna", indicato in sanscrito anche come tetralemma — è la struttura fondamentale dell'argomentazione nagarjuniana. Per qualsiasi proposizione P e per qualsiasi oggetto di cui si predica P , Nagarjuna distingue quattro posizioni: (1) P è vero; (2) P è falso ($\neg P$ è vero); (3) P è sia vero che falso ($P \wedge \neg P$); (4) P non è né vero né falso ($\neg P \wedge \neg \neg P$). Nel pensiero di Nagarjuna, applicato sistematicamente alle proposizioni fondamentali della metafisica e dell'ontologia buddhista, tutte e quattro le posizioni del catuskoti vengono negate per ogni oggetto reale: nessuna di esse cattura adeguatamente la natura dei fenomeni, che è *śūnyatā* — vacuità, mancanza di natura intrinseca.

La struttura formale del catuskoti è stata analizzata da Graham Priest e dalla logica della dialettica nel contesto delle logiche paraconsistenti: il tetralemma buddhista è un sistema a quattro valori di verità $\{T, F, B, N\}$ (vero, falso, entrambi, nessuno) che corrisponde alla logica di Belnap FOUR — il sistema logico a quattro valori sviluppato da Nuel Belnap per i database con informazioni potenzialmente contraddittorie. In questa logica:

Valori: $\{ T \text{ (solo vero)}, F \text{ (solo falso)}, B \text{ (entrambi)}, N \text{ (nessuno)} \}$
 $\neg T = F, \neg F = T, \neg B = B, \neg N = N$
 $T \wedge B = B, F \vee N = N, \text{ ecc.}$

La negazione dei valori estremi (T e F) è quella classica; la negazione dei valori misti (B e N) è involutiva — il sia-vero-che-falso negato rimane sia-vero-che-falso; il né-vero-né-falso negato rimane né-vero-né-falso. Questa struttura formale corrisponde alla distinzione nagarjuniana tra le proposizioni ordinarie — per le quali la negazione classica funziona — e le proposizioni metafisiche fondamentali — per le quali la negazione non produce una

transizione verso il polo opposto ma una stabilizzazione nella posizione mista o nella posizione vuota.

Per il cervello, questa struttura ha una traduzione nel linguaggio degli attrattori: le proposizioni ordinarie hanno due attrattori stabili (vero e falso), tra i quali la negazione produce transizione; le proposizioni metafisicamente fondamentali hanno attrattori degeneri o piatti — il sistema non converge stabilmente verso nessun polo, e la negazione non produce transizione ma oscillazione stabile tra i poli. Il catuskoti è la logica del paesaggio energetico degenerato: la logica dei sistemi in cui non c'è un attrattore profondo e il sistema rimane in uno stato di bassa energia libera distribuita uniformemente tra le possibilità.

La prasanga — il metodo argomentativo nagarjuniano di *reductio ad absurdum* sistematica, attraverso cui ogni posizione metafisica viene ridotta alla sua propria contraddizione — non è la *reductio* classica della logica occidentale: non cerca di confutare una tesi derivando da essa una contraddizione e poi affermando il contrario. Cerca piuttosto di mostrare che qualsiasi posizione determinata — qualsiasi tentativo di affermare o negare qualcosa di fondamentale sulla realtà — genera inevitabilmente contraddizioni, e che quindi la posizione corretta non è nessuna delle quattro corna del tetralemma ma il silenzio che si trova al di là di tutte e quattro. Il silenzio non come negazione di ogni posizione — che sarebbe già una posizione — ma come abbandono dell'intera struttura di affermazione e negazione.

Neurobiologicamente, questo silenzio nagarjuniano può essere avvicinato — con tutta la cautela richiesta dal salto disciplinare — allo stato cognitivo prodotto da pratiche contemplative intensive che riducono sistematicamente l'attività del Default Mode Network e delle reti di elaborazione proposizionale. Gli studi di neuroimaging sulle pratiche meditative avanzate mostrano una riduzione dell'attività nelle aree associate alla Teoria della Mente, al self-referential processing, e all'elaborazione delle proposizioni su sé e sul mondo — le aree che implementano, nel framework qui sviluppato, la struttura dual-coding della negazione proposizionale. Il silenzio meditativo è, in questa lettura, non la negazione di tutte le proposizioni ma la disattivazione dei sistemi che producono la struttura posizionale della coscienza — e quindi la dissoluzione della negazione non perché sia stata affermata, ma perché il substrato che la genera ha temporaneamente ridotto la propria attività al di sotto della soglia necessaria per l'elaborazione proposizionale.

LV. La proposta di una teoria unificata: la negazione come struttura topologica del modello generativo

Dopo aver attraversato logica formale, neuroscienza cognitiva, fenomenologia, psicologia dello sviluppo, filogenetica, teoria matematica dei topos, pragmatica, logica temporale, principio di energia libera e filosofia comparata, è possibile tentare una sintesi parziale che non pretenda di risolvere il problema della

negazione — che, come si è argomentato nella conclusione del saggio originale, non è il tipo di problema che si risolve — ma di articolare la struttura del campo che lo contiene.

La tesi unificante che emerge dall'analisi è la seguente: la negazione non è una singola operazione cognitiva ma una famiglia di operazioni strutturalmente correlate che condividono una proprietà fondamentale, la quale può essere espressa topologicamente. Questa proprietà è la costruzione di un confine — in senso topologico, una struttura di separazione — all'interno dello spazio degli stati del modello generativo. La negazione di P è la costruzione di un confine che separa P dal resto dello spazio degli stati: colloca P in una regione separata, marcata come inaccessibile o non-attuale rispetto alla regione in cui si trova il modello del mondo corrente.

Questo confine ha diverse proprietà topologiche a seconda del tipo di negazione:

Per la **negazione accidentale** ($\neg P$ dove P è contingentemente falsa), il confine è permeabile: c'è un percorso continuo nello spazio degli stati che porta dal modello attuale alla regione di P attraverso una sequenza di aggiornamenti bayesiani. Il confine è una barriera di potenziale che può essere superata con sufficiente evidenza.

Per la **negazione necessaria** ($\Box \neg P$ dove P è strutturalmente impossibile), il confine è impermeabile: non esiste percorso continuo nello spazio degli stati del modello che porti nella regione di P . Topologicamente, P appartiene a una componente disconnessa dello spazio degli stati — una regione inaccessibile non per la forza della barriera ma per la struttura topologica dello spazio.

Per la **negazione temporale futura** ($\neg F P$ dove una traiettoria viene chiusa), il confine è asimmetrico rispetto al tempo: la regione di P era accessibile prima della chiusura, è inaccessibile dopo. Il confine non è nello spazio degli stati ma nello spazio-tempo del modello generativo.

Per la **negazione paraconsistente** (situazioni di alta credenza sia in P che in $\neg P$), il confine è sfumato: le due regioni si sovrappongono in una zona di confine dove sia P che $\neg P$ hanno significativa massa di probabilità. Topologicamente, P e $\neg P$ non sono regionicomplementari con confine netto, ma regioni parzialmente sovrapposte con transizione graduale.

Per la **negazione autoreferenziale** ($\neg P$ dove P parla di se stessa), il confine è non-orientabile: tentare di determinare da quale lato del confine ci si trova produce una struttura analoga alla striscia di Möbius, in cui l'interno e l'esterno sono la stessa superficie. Il confine non divide lo spazio degli stati in due regioni separate ma produce una singolarità topologica.

Questa caratterizzazione topologica della negazione — come costruzione di confini di diverse strutture geometriche nello spazio degli stati del modello

generativo — è la sintesi formale che emerge dall'intera analisi. Essa permette di trattare tutti i tipi di negazione come istanze di una struttura matematica unica (la geometria differenziale degli spazi degli stati e le topologie dei loro confini) pur mantenendo le distinzioni funzionali e computazionali tra i tipi che l'analisi empirica ha rivelato.

Il substrato biologico di questa struttura topologica è il paesaggio energetico del sistema corticale — la distribuzione dell'energia libera nello spazio degli stati del modello generativo — la cui topologia è determinata dalla storia dell'apprendimento del sistema, dalla struttura delle sue connessioni sinaptiche, e dalla dinamica dei suoi prior. La negazione, in questa prospettiva finale, è la manipolazione della topologia del paesaggio energetico: la creazione, la modifica o l'annullamento di confini che separano regioni dello spazio degli stati. E il cervello umano che pensa la negazione è il sistema biologico che, attraverso millenni di evoluzione e anni di sviluppo individuale, ha acquisito la capacità più sofisticata di modificare dinamicamente la propria topologia interna — di costruire e abbattere confini nello spazio delle proprie possibilità rappresentazionali — producendo così la struttura della mente come spazio aperto sulla propria assenza.

LVI. Il residuo irriducibile come problema aperto: fenomenologia, coscienza e la negazione della negazione

Ogni sintesi onesta deve confrontarsi con ciò che sfugge alla sintesi. Tre residui irriducibili restano al margine del framework teorico sviluppato, e il loro riconoscimento esplicito è più prezioso di qualsiasi tentativo di dissolverli prematuramente.

Il primo residuo è quello già individuato nel saggio originale: il carattere qualitativo dell'esperienza del vuoto intenzionale. Nessuna descrizione formale della negazione come struttura topologica del paesaggio energetico cattura ciò che significa per un soggetto esperire l'assenza come tale — il peculiare fenomeno di presenza dell'assente che caratterizza la cognizione umana a differenza di qualsiasi sistema puramente computazionale. Questo residuo non è un problema tecnico in attesa di soluzione: è il confine del problema difficile della coscienza applicato al caso specifico della negazione, e come tale è strutturalmente resistente a qualsiasi spiegazione funzionale.

Il secondo residuo riguarda la normativa della negazione: il fatto che alcune negazioni siano corrette e altre sbagliate, che ci sia una differenza tra il negare giustificato e il negare arbitrario. Il framework della minimizzazione dell'energia libera descrive come il sistema neghi, non come dovrebbe negare. La normativa della negazione — il fondamento della sua correttezza epistemica — richiede una teoria della razionalità che nessuna teoria puramente meccanicistica può fornire dall'interno: richiede di uscire dal registro descrittivo e entrare nel registro normativo, il che significa confrontarsi con le tradizioni dell'epistemologia e della

logica non come fonti di strutture formali da tradurre in neuroscienze, ma come riflessioni autonome sulle condizioni di possibilità della conoscenza corretta.

Il terzo residuo è il più sottile e forse il più profondo: la negazione della negazione stessa. Tutto il saggio ha trattato la negazione come una struttura cognitiva che il cervello implementa, descrive e usa. Ma c'è una domanda che il saggio non ha formulato e che sfugge al suo impianto metodologico: è possibile negare la struttura della negazione stessa? È possibile un pensiero che non sia strutturato dalla distinzione tra affermazione e negazione — non un pensiero privo di contenuto, ma un pensiero che abiti un'esperienza del reale non organizzata intorno a questa distinzione?

Nagarjuna rispondeva di sì, e chiamava questa possibilità il silenzio del Buddha. La fisica quantistica suggerisce, nella sua interpretazione più radicale, che a certi livelli di realtà le proprietà definite — e quindi la distinzione tra il loro possedere e non-possedere — non sono applicabili prima della misura. La meditazione contemplativa avanzata riporta stati di coscienza in cui la struttura soggetto-oggetto — e con essa la struttura affermazione-negazione — si dissolve temporaneamente senza che il soggetto perda coscienza.

Questi tre campi — filosofia buddhista, fisica quantistica, neuroscienze contemplative — convergono nell'indicazione che la negazione non è la struttura ultima della mente, ma una struttura derivata che emerge da un livello più profondo in cui la distinzione tra essere e non-essere non è ancora costituita. Questo livello più profondo — chiamato *śūnyatā* dai buddhisti, stato del quantum field dai fisici, consapevolezza pura dai contemplativi — non è accessibile alla neuroscienza attraverso i metodi standard, perché questi metodi presuppongono già la struttura soggetto-oggetto e affermazione-negazione che si tratta di mettere in questione.

La negazione, dunque, non è soltanto un problema della neuroscienza cognitiva: è, al suo confine più radicale, un problema della filosofia nel suo senso più antico — il problema di ciò che precede la struttura stessa del pensiero concettuale, e di come questo precedente possa essere, se mai può esserlo, detto.

Il programma sperimentale — previsioni, paradigmi e misure

LVII. La necessità di un programma sperimentale esplicito

Un'analisi teorica della negazione che ambisca a dialogare con la neuroscienza empirica non può limitarsi a citare studi esistenti come conferme parziali di una struttura concettuale costruita a posteriori: deve formulare previsioni che la struttura teorica genera in anticipo rispetto agli esperimenti, previsioni che abbiano la proprietà di falsificabilità nel senso popperiano — che siano cioè tali da poter essere confutate da risultati sperimentali determinati, e non soltanto

confermate da qualunque esito. Questo passaggio dal registro descrittivo-interpretativo al registro predittivo-falsificabile è il salto epistemologico che trasforma un framework teorico in una teoria scientifica nel senso pieno del termine.

Il framework sviluppato nelle sezioni precedenti genera un numero considerevole di previsioni sperimentali. Non tutte sono ugualmente testate con le tecnologie correnti: alcune richiedono paradigmi già disponibili ma non ancora applicati al problema della negazione; altre richiedono sviluppi metodologici che sono sulla frontiera della neuroscienza sperimentale ma non ancora di routine. Questa differenza di accessibilità sperimentale è rilevante: le previsioni più immediatamente testabili hanno maggior valore diagnostico a breve termine, anche se le previsioni più profonde riguardano dimensioni del problema che le misure correnti non raggiungono.

Si organizzano le previsioni per dominio di misura, specificando in ciascun caso la previsione precisa, il paradigma sperimentale che la testa, la variabile dipendente critica, e il risultato alternativo che falsificherebbe la previsione.

LVIII. Previsioni per l'elettrofisiologia (ERP e oscillazioni)

La componente ERP più studiata in relazione alla negazione è la N400 — la negatività centrale e posteriore con picco intorno a 400 millisecondi dallo stimolo critico, originariamente associata all'anomalia semantica e successivamente collegata a qualsiasi discrepanza tra l'input linguistico e le aspettative del sistema. La N400 è un misuratore dell'errore di previsione semantico: la sua ampiezza riflette la grandezza della discrepanza tra il contesto precedente e la parola ricevuta, pesata per la precisione delle previsioni del sistema.

La **prima previsione** riguarda la modulazione della N400 dalla profondità concettuale della negazione. Il framework propone che la negazione necessaria — la negazione di proprietà definitorie di un concetto — produca una N400 di struttura temporale e topografica diversa dalla negazione contingente. Specificamente: la negazione di una proprietà necessaria (la frase "il quadrato ha tre lati" seguita dal correttivo "no, quattro") dovrebbe produrre una N400 con distribuzione più anteriore e latenza leggermente più tardiva rispetto alla negazione di una proprietà contingente ("Mario era a Roma ieri" seguito da "no, era a Milano"), riflettendo il maggiore coinvolgimento delle cortecce prefrontali nel processing della violazione definitoria rispetto alla semplice violazione contestuale. Il risultato alternativo che falsificherebbe questa previsione sarebbe l'assenza di differenza nella morfologia della N400 tra i due tipi di negazione — un risultato che suggerirebbe che il sistema elettrofisiologico non distingue la negazione necessaria da quella contingente al livello della risposta semantica rapida.

La **seconda previsione** riguarda la componente LFN (Late Frontal Negativity) e la sua sensibilità alla struttura del dual-coding. Se la negazione proposizionale richiede la costruzione della rappresentazione affermativa prima della sua modificazione negativa — come argomentato seguendo Husserl e come suggerito dagli studi di compatibilità azione-parola — allora la LFN dovrebbe avere una struttura temporale bifasica: un picco precoce (200–300 ms) associato alla costruzione della rappresentazione affermativa, e un picco tardivo (400–600 ms) associato alla sua marcatura negativa. Questa struttura bifasica non è stata cercata sistematicamente nella letteratura esistente, che ha tipicamente analizzato la LFN come componente unitaria. Il paradigma sperimentale richiede la separazione temporale delle due fasi attraverso la manipolazione della velocità di presentazione: presentazioni più lente (con stimolo onset asynchrony aumentato) dovrebbero aumentare il gap tra i due picchi e renderli più separabili nell'analisi time-frequency.

La **terza previsione** riguarda le oscillazioni gamma e la loro relazione con il dual-coding della negazione. Nel framework della codifica predittiva, il segnale gamma (30–80 Hz) è il vettore degli errori di previsione ascendenti — il segnale che comunica la discrepanza tra input e previsione. Se la negazione proposizionale produce simultaneamente un modello attuale (alta precisione, basso errore di previsione) e un modello controfattuale (bassa precisione, costruzione simulativa), si prevede una desincronizzazione gamma localizzata nelle aree che elaborano il contenuto specifico della proposizione negata (aree visive per contenuti visuali, aree motorie per contenuti motori) accompagnata da una sincronizzazione gamma nelle regioni prefrontali associate al reality monitoring. Questa dissociazione spaziale — desincronizzazione posteriore, sincronizzazione anteriore — è analoga al pattern osservato nei compiti di soppressione cognitiva, ma la previsione specifica della sincronizzazione prefrontale come correlato della marcatura negativa del modello controfattuale non è mai stata testata direttamente nel contesto della negazione proposizionale.

La **quarta previsione**, forse la più diagnosticamente potente, riguarda le oscillazioni theta ippocampali durante la comprensione di negazioni controfattuali forti. L'ippocampo, come sistema di simulazione del non-attuale, genera oscillazioni theta (4–8 Hz) durante i processi di ricombinazione flessibile degli elementi esperienziali in configurazioni nuove — i processi che si è proposto stiano alla base della simulazione controfattuale richiesta dalla negazione proposizionale matura. La previsione è che le frasi che implicano scenari controfattuali ricchi e specifici ("Mario non è andato a Berlino, dove avrebbe potuto visitare il Reichstag") producano maggiore potenza theta ippocampale rispetto a frasi con negazioni strutturalmente analoghe ma contenuto controfattuale povero o impossibile ("il numero 7 non è pari"). Questa previsione richiede misure intracraniali o MEG con localizzazione ippocampale adeguata — tecnicamente accessibili nei contesti di ricerca con pazienti epilettici e con magnetoencefalografi ad alta densità.

LIX. Previsioni per il neuroimaging funzionale (fMRI e PET)

Il neuroimaging funzionale con risonanza magnetica offre la risoluzione spaziale necessaria per distinguere le attivazioni nelle diverse sottoregioni prefrontali e temporali coinvolte nel processing della negazione, anche se la sua risoluzione temporale — nell'ordine dei secondi, non dei millisecondi — non permette di catturare la struttura temporale fine del processo.

La **prima previsione** per l'fMRI riguarda la dissociazione tra vmPFC e dlPFC nella elaborazione della negazione descrittiva versus inferenziale. La distinzione proposta nella sezione IX del saggio originale tra negazione descrittiva (che tende alla soppressione della rappresentazione negata) e negazione inferenziale (che tende al dual-coding) dovrebbe produrre pattern di attivazione differenziali: la negazione descrittiva dovrebbe coinvolgere prevalentemente la dlPFC come sistema di aggiornamento del modello corrente, con relativa disattivazione delle aree del Default Mode Network associate alla simulazione; la negazione inferenziale dovrebbe coinvolgere in modo bilanciato sia la dlPFC sia le aree DMN — in particolare l'ippocampo e la corteccia cingolata posteriore — riflettendo il mantenimento attivo del modello controfattuale richiesto dal ragionamento ipotetico. Il paradigma sperimentale dovrebbe presentare coppie di frasi negative con diverso contesto sintattico-pragmatico che determina la lettura descrittiva o inferenziale ("la mela non è rossa" in contesto descrittivo vs. "se la mela non fosse rossa..." in contesto inferenziale).

La **seconda previsione** per l'fMRI riguarda la connettività funzionale tra ippocampo e vmPFC durante la comprensione di negazioni con diverso grado di accessibilità controfattuale. Se la negazione necessaria non richiede simulazione ippocampale mentre la negazione accidentale la richiede, la connettività funzionale ippocampo-vmPFC dovrebbe essere significativamente maggiore durante la comprensione di negazioni accidentali ricche ("Mario non è andato al cinema, sarebbe stato divertente") rispetto alle negazioni necessarie ("un triangolo non ha quattro angoli"). Questa previsione è verificabile con la Dynamic Causal Modelling, che permette di stimare le direzioni delle influenze causali tra regioni dalla struttura del segnale BOLD.

La **terza previsione** riguarda il reality monitoring e la sua modulazione con l'età. Se la capacità di mantenere il differenziale di precisione tra modello attuale e modello controfattuale declina con l'invecchiamento cognitivo — come ci si aspetta sulla base della vulnerabilità delle cortecce prefrontali mediali all'atrofia età-correlata — allora la negazione proposizionale negli anziani cognitivamente sani dovrebbe mostrare maggiore attivazione vmPFC (per compensare la riduzione di efficienza del reality monitoring) e tempi di reazione più lunghi nella verifica di frasi negative, con una dissociazione tra le negazioni necessarie (relativamente preservate, perché non dipendono dalla simulazione controfattuale ma dalla struttura del modello definitório) e le negazioni accidentali ricche (più compromesse, perché dipendono dal circuito

ippocampo-vmPFC). Questa previsione è verificabile in studi di neuroimaging su popolazioni di adulti sani di diverse età.

LX. Previsioni per i paradigmi comportamentali

I paradigmi comportamentali — tempi di reazione, accuratezza, misure di compatibilità azione-parola, paradigmi di priming semantico — offrono la via sperimentale più immediatamente accessibile per testare alcune delle previsioni più specifiche del framework.

La **prima previsione comportamentale** riguarda l'asimmetria tra il costo computazionale della negazione necessaria e di quella accidentale nei compiti di verifica di frasi. La previsione è che la negazione necessaria produca tempi di risposta più brevi della negazione accidentale in compiti di verifica rapida della verità/falsità di frasi semplici, nonostante la sua maggiore complessità logica. La ragione è che la negazione necessaria non richiede la costruzione della simulazione controfattuale ippocampale — il sistema può produrre direttamente il segnale di incoerenza strutturale senza passare per la costruzione del mondo alternativo — mentre la negazione accidentale richiede questa costruzione intermedia. Il risultato alternativo falsificante sarebbe la sistematica maggior lentezza delle negazioni necessarie, che suggerirebbe che la profondità concettuale del contenuto negato domina sul costo computazionale della simulazione controfattuale.

La **seconda previsione comportamentale** riguarda la struttura degli errori nella comprensione di negazioni annidate in funzione del carico metacognitivo. Le negazioni annidate — "non è falso che non sia impossibile che P" — richiedono il tracking di un numero crescente di inversioni del valore di verità, e i soggetti commettono errori sistematici che tendono verso l'affermazione della proposizione con il minor numero di negazioni. La previsione del framework è che questo errore sia modulato dalla salienza del contenuto proposizionale: quando P descrive un contenuto concretamente simulabile, gli errori di tracking delle negazioni annidate dovrebbero essere maggiori perché il sistema è attirato dalla simulazione del contenuto piuttosto che dal tracking formale delle inversioni; quando P è astratta e difficilmente simulabile, il tracking formale delle negazioni dovrebbe essere più accurato perché il sistema non è distratto dalla simulazione controfattuale. Questa interazione — salienza del contenuto $\times$ numero di negazioni annidate $\times$ accuratezza — è una previsione precisa che nessuna teoria puramente sintattica della negazione può generare.

La **terza previsione comportamentale** riguarda la compatibilità motoria della negazione in funzione del contesto temporale. Gli studi di Glenberg e Bergen hanno mostrato che la comprensione di frasi che descrivono azioni produce facilitazione o interferenza con azioni motorie compatibili. La previsione per la negazione è che questa compatibilità motoria sia modulata dalla struttura temporale della frase negativa: "non afferrare la tazza" dovrebbe produrre interferenza motoria con l'azione di afferrare (perché il contenuto motorio viene

costruito prima della negazione), ma questa interferenza dovrebbe essere significativamente ridotta se la negazione è temporalmente lontana rispetto al contesto motorio immediato ("quando entrerai nella stanza non afferrare la tazza"). La ragione è che la negazione a distanza temporale permette al sistema di costruire la rappresentazione del futuro senza attivare il sistema motorio corrente, mentre la negazione immediata arriva troppo tardi per impedire l'attivazione motoria prodotta dalla simulazione del contenuto.

La negazione come atto linguistico — illocuzione, perlocuzione e le condizioni del negare

LXI. Austin, Searle e la negazione come atto

La teoria degli atti linguistici di John Langshaw Austin — formulata nelle William James Lectures del 1955, pubblicate postume in *How to Do Things with Words* (1962) e sviluppata da John Searle in *Speech Acts* (1969) e in *Expression and Meaning* (1979) — ha trasformato la filosofia del linguaggio spostando il fuoco dall'analisi delle condizioni di verità delle proposizioni all'analisi delle azioni che gli esseri umani compiono attraverso l'uso del linguaggio. La distinzione austiniana fondamentale tra atto locutorio (il dire qualcosa con un certo senso e riferimento), atto illocutorio (il compiere un'azione attraverso il dire, come promettere, ordinare, dichiarare) e atto perlocutorio (il produrre un effetto nell'interlocutore attraverso il dire, come persuadere, intimidire, consolare) apre alla negazione una prospettiva che la logica proposizionale non può offrire.

La negazione, nella teoria degli atti linguistici, non è soltanto un operatore che inverte il valore di verità di una proposizione: è una famiglia di atti illocutori distinti che condividono il contenuto proposizionale negativo ma differiscono per la loro forza illocutoria e per le loro condizioni di felicità. Negare, rifiutare, smentire, contraddire, disconoscere, sconfessare, revocare — questi non sono sinonimi della stessa operazione logica, ma atti linguistici categorialmente diversi con strutture pragmatiche e condizioni di appropriatezza radicalmente differenti.

Questa molteplicità degli atti negativi è stata in larga misura ignorata dalla teoria linguistica della negazione, che ha privilegiato la dimensione semantico-proposizionale. Ma ha conseguenze neurologiche dirette: se diversi atti negativi coinvolgono strutture pragmatiche diverse, ci si aspetta che i loro correlati neurali differiscano sistematicamente — non soltanto nei circuiti di elaborazione del linguaggio, ma nei circuiti di pianificazione dell'azione comunicativa e nei sistemi di Teoria della Mente che modellano l'effetto atteso sull'interlocutore.

LXII. La tassonomia degli atti negativi e le loro condizioni di felicità

Sviluppare una tassonomia precisa degli atti negativi richiede di distinguere almeno cinque dimensioni che intersecano in modi diversi per i diversi atti:

La prima dimensione è l'oggetto della negazione: si può negare un fatto (rifiuto assertivo), un'azione proposta (rifiuto commissivo), un'autorizzazione (revoca), un'identità o una descrizione (disconoscimento), una relazione (ripudio). Ciascuno di questi oggetti definisce un diverso dominio di applicazione dell'atto negativo.

La seconda dimensione è il destinatario: la negazione può essere rivolta all'interlocutore diretto (smentita), a terzi presenti (testimonianza negativa), a se stessi (auto-correzione, ritrattazione), o a un pubblico indefinito (sconfessione pubblica). La struttura della Teoria della Mente richiesta varia sistematicamente con il numero e il tipo di destinatari.

La terza dimensione è il tempo della proposizione negata: si può negare il passato (smentita storica), il presente (negazione descrittiva), il futuro (rifiuto prospettico), o un'eventualità ipotetica (negazione controfattuale preventiva). Il correlato neurobiologico di questa dimensione è la struttura temporale del modello generativo coinvolto, come analizzato nella Parte Nona.

La quarta dimensione è la forza della negazione: da una negazione debole che esprime dubbio o incertezza ("non sono sicuro che sia così") a una negazione forte che esprime certezza epistemica ("è assolutamente falso") a una negazione performativa che produce effetti istituzionali indipendentemente dalla propria sincerità ("dichiaro nullo e privo di effetti").

La quinta dimensione è lo statuto del negante rispetto alla proposizione negata: si può negare qualcosa che si sapeva precedentemente vero (ritrattazione), qualcosa che non si è mai creduto (smentita esterna), qualcosa che si credeva vero ma si ha ora evidenza sia falso (correzione), qualcosa che l'interlocutore crede ma che non si condivide (disaccordo), qualcosa che nessuno ha esplicitamente affermato ma che la struttura della conversazione presuppone (negazione presuppositiva).

Questa tassonomia a cinque dimensioni genera uno spazio di atti negativi potenzialmente molto ricco, ma non tutti i punti dello spazio sono ugualmente grammaticali e pragmaticamente coerenti: ci sono vincoli di combinazione che riflettono sia la struttura logica delle condizioni di felicità sia la struttura socio-istituzionale dei contesti in cui questi atti si compiono. Le condizioni di felicità di una smentita, per esempio, includono che il negante abbia accesso privilegiato alla verità della proposizione negata (o lo rivendichi), che l'interlocutore abbia affermato o implicitamente sostenuto la proposizione, e che il contesto sia appropriato per la contestazione delle credenze altrui. Se queste condizioni non sono soddisfatte, la smentita non è falsa o vera — è infelice nel senso austiniano: non raggiunge lo statuto di atto compiuto.

LXIII. La negazione performativa e il suo paradosso

Tra gli atti negativi, quello che pone il problema teorico più acuto è la negazione performativa: l'atto linguistico che non descrive o valuta uno stato di fatto ma lo produce attraverso la sua stessa enunciazione. "Dichiaro nullo il contratto", "Ti scomunico", "Revoco il mio consenso": questi atti non descrivono una nullità, una scomunica, una revoca preesistenti — le producono attraverso l'atto stesso del dire, nelle circostanze appropriate e da parte di un soggetto con l'autorità istituzionale richiesta.

La negazione performativa pone un problema che la logica proposizionale non può formulare: non è la negazione di una proposizione, ma la negazione di uno stato istituzionale-sociale che ha esistenza soltanto in quanto è collettivamente riconosciuto. La nullità giuridica di un contratto non è un fatto naturale come il colore di una mela: è una proprietà che esiste nel senso di John Searle — come fatto istituzionale, dipendente dall'accordo collettivo circa le regole che le attribuiscono esistenza. Negare performativamente questo stato significa produrre un nuovo fatto istituzionale che cancella il precedente — non attraverso un'azione sul mondo fisico, ma attraverso un'azione sul mondo sociale.

Neurologicamente, la comprensione e la produzione di negazioni performative richiedono strutture cognitive che vanno ben al di là di quelle coinvolte nella negazione proposizionale ordinaria: richiedono la rappresentazione delle strutture istituzionali e del loro funzionamento (il contratto come fatto istituzionale, l'autorità come proprietà conferita da regole sociali), la simulazione delle conseguenze sociali dell'atto negativo (che cosa cambia nel mondo istituzionale dopo la revoca), e la valutazione dell'autorità del negante rispetto all'atto che compie. Queste operazioni coinvolgono la corteccia prefrontale ventromediale nel suo ruolo di integrazione tra conoscenza sociale e valutazione delle conseguenze, le regioni temporali anteriori nella rappresentazione semantica delle strutture socio-istituzionali, e la giunzione temporo-parietale nella simulazione della prospettiva dell'interlocutore.

Il paradosso della negazione performativa è che essa sembra contraddire la struttura husserliana della posizione-e-cancellazione: non costruisce prima la rappresentazione dello stato negato e poi la cancella — lo cancella direttamente attraverso l'azione. Ma questa contraddizione è soltanto apparente: l'atto performativo presuppone la rappresentazione dello stato che cancella (per poterlo cancellare, il negante deve rappresentarselo come esistente), e la cancellazione avviene non nella rappresentazione del soggetto ma nel mondo istituzionale condiviso. La struttura husserliana si conserva al livello del soggetto individuale; la novità della negazione performativa è che essa produce effetti nel mondo sociale che non si riducono all'aggiornamento del modello del singolo agente.

LXIV. La negazione sincera e quella insincera: menzogna, ironia e reticenza

Austin aveva sottolineato che le condizioni di felicità degli atti linguistici includono una condizione di sincerità: chi compie l'atto deve trovarsi nello stato psicologico appropriato. Chi promette sinceramente ha l'intenzione di mantenere la promessa; chi si scusa sinceramente prova rincrescimento per l'offesa. Applicata alla negazione assertiva, la condizione di sincerità richiede che il negante creda effettivamente che la proposizione negata sia falsa.

La violazione deliberata di questa condizione di sincerità — la negazione di ciò che si sa vero — è la menzogna, una delle forme di comportamento linguistico più studiate e al tempo stesso meno comprese nella sua struttura cognitiva. La menzogna non è soltanto la produzione di un enunciato falso: è la produzione di un enunciato falso con l'intenzione di produrre nell'interlocutore una credenza falsa, e con la consapevolezza che si sta compiendo questa operazione. Questa struttura triplice — falsità dell'enunciato, intenzione di ingannare, consapevolezza dell'intenzione — richiede una Teoria della Mente di secondo ordine e una capacità di mantenere simultaneamente la propria credenza vera e la rappresentazione della credenza falsa che si vuole indurre nell'altro.

Neurologicamente, la produzione di negazioni menzognere è stata associata in numerosi studi a una maggiore attivazione della corteccia prefrontale dorsolaterale e della corteccia cingolata anteriore rispetto alla produzione di negazioni veritiere — un pattern interpretato come riflesso del maggiore sforzo cognitivo richiesto per inibire la risposta veritiera e sostituirla con la risposta falsa. Questa interpretazione è coerente con il framework proposto: la menzogna richiede la costruzione del modello del mondo reale (con la credenza vera), la costruzione del modello alternativo da comunicare (con la credenza falsa che si vuole indurre), e il mantenimento della separazione tra i due modelli — esattamente la struttura del dual-coding della negazione, ma applicata non alla distinzione reale/controfattuale bensì alla distinzione tra il proprio modello e il modello da trasmettere all'interlocutore.

La violazione della condizione di sincerità nella direzione opposta — la negazione di ciò che si crede vero non per ingannare ma per comunicare indirettamente la verità — è l'ironia. "Non è proprio il caso di dire che la tua relazione fosse impeccabile" comunica, attraverso la negazione di una litote, la valutazione molto negativa della relazione — e lo fa in un modo che presuppone nell'interlocutore la capacità di inferire il contrario di quanto letteralmente detto. L'ironia è la forma di negazione pragmaticamente più sofisticata: richiede che sia il parlante sia l'ascoltatore rappresentino l'enunciato letterale come non-sincero, e che questa non-sincerità sia mutuamente manifesta — entrambi sanno che entrambi sanno che il significato inteso è l'opposto del significato letterale.

Neurologicamente, l'elaborazione dell'ironia ha un profilo fMRI distinto dall'elaborazione della negazione letterale, con maggiore coinvolgimento della corteccia prefrontale mediale e della giunzione temporo-parietale destra — le stesse regioni associate alla Teoria della Mente e alla simulazione della prospettiva altrui. Questa differenza è coerente con la struttura pragmatica dell'ironia come

negazione di secondo ordine: non soltanto la negazione di un contenuto, ma la negazione della sincerità dell'enunciato che porta quel contenuto — un'operazione che richiede di uscire dal livello proposizionale per salire al livello della relazione tra parlante, enunciato e intenzione comunicativa.

La reticenza — il non-dire che equivale a una forma di negazione implicita, il silenzio eloquente che comunica negazione senza formularla — è forse la forma più sottile e neuroscientificamente meno esplorata degli atti negativi. "Non rispondere è già una risposta", dice la massima popolare; e in effetti il silenzio in risposta a una domanda diretta comunica, per implicatura conversazionale griceana, la negazione di ciò che sarebbe stato l'esito atteso della risposta. La reticenza è una negazione performativa implicita che sfrutta la struttura delle aspettative conversazionali per produrre il contenuto negativo senza enunciarlo. Il suo correlato neurobiologico — la soppressione della risposta linguistica che sarebbe stata prodotta in assenza dell'intenzione reticente — coinvolge i sistemi di inibizione della risposta motoria (il circuito cortico-striatale-talamo-corticale associato all'inibizione dell'azione) applicati al dominio della pianificazione linguistica, producendo un'inibizione dell'output che è, strutturalmente, la forma più diretta di negazione come inibizione: non la negazione di una proposizione, ma la negazione dell'atto stesso di affermare.

LXV. La forza illocutoria della negazione e la sua modulazione prosodica

La forza illocutoria di un atto negativo — la distinzione tra una negazione assertiva debole e una negazione assertiva forte, tra il disaccordo cortese e la smentita diretta, tra il rifiuto tentativo e il rifiuto definitivo — è codificata in larga misura non nella struttura proposizionale ma nella prosodia: il tono, il ritmo, l'intensità, la durata delle vocali, la struttura dell'intonazione. Questa codifica prosodica della forza illocutoria è universale — tutte le lingue studiate usano la prosodia per distinguere le diverse sfumature degli atti negativi — e ha correlati neurobiologici nella percezione e produzione del parlato che permettono di connettere la teoria degli atti linguistici con la neuroscienza del linguaggio.

Le corteccie uditive superiori, in particolare la corteccia di Heschl e le regioni del planum temporale nell'emisfero destro, sono coinvolte nella elaborazione dei pattern prosodici dell'intonazione emotiva e pragmatica. Studi di ERP con la componente CPS (Closure Positive Shift) — la positività centroparietale associata ai confini intonativi — hanno mostrato che il cervello elabora la struttura prosodica in parallelo con quella sintattica e semantica, integrandole in una rappresentazione pragmatica unificata dell'atto linguistico nel giro di 600–800 millisecondi dall'inizio della frase.

Per la negazione, la previsione specifica è che la forza illocutoria dell'atto negativo — codificata prosodicamente — moduli l'elaborazione del contenuto

proposizionale negato: una negazione prosodicamente forte (tono discendente brusco, intensità aumentata, durata ridotta della vocale del morfema negativo) dovrebbe sopprimere più rapidamente la simulazione del contenuto negato rispetto a una negazione prosodicamente debole (tono più sostenuto, intensità normale, durata vocale preservata). Questa previsione, testabile attraverso la manipolazione sistematica dei pattern prosodici a contenuto proposizionale costante, permetterebbe di dimostrare che la forza illocutoria — una dimensione pragmatica dell'atto — non è elaborata separatamente dal contenuto proposizionale ma modula in tempo reale la struttura computazionale della negazione.

LXVI. La negazione nei discorsi istituzionali: diritto, medicina, politica

Gli atti negativi hanno una rilevanza pratica particolarmente elevata nei discorsi istituzionali — il diritto, la medicina, la politica — perché in questi contesti le negazioni producono effetti nel mondo che vanno ben al di là dell'aggiornamento delle credenze degli interlocutori. Una sentenza che nega la colpevolezza di un imputato non aggiorna soltanto le credenze del tribunale: produce uno stato giuridico con conseguenze pratiche che modificano la struttura del mondo sociale. Una diagnosi che nega la presenza di una malattia non aggiorna soltanto le credenze del medico: determina percorsi terapeutici, stati assicurativi, scelte di vita.

In questi contesti, la struttura cognitiva della negazione si intreccia con strutture istituzionali che ne moltiplicano la complessità in modo che le teorie puramente proposizionali non catturano. La negazione giuridica — "non colpevole" nel sistema anglosassone, "assolto" nei sistemi di diritto civile — è interessante perché nei sistemi di common law il verdetto "not guilty" non equivale all'affermazione "innocent": è la negazione della colpevolezza oltre ogni ragionevole dubbio, che lascia aperto lo spazio epistemico della colpevolezza possibile pur producendo gli effetti giuridici dell'innocenza. Questa struttura — la negazione che produce stati istituzionali definiti senza produrre certezza epistemica — è l'inverso della negazione necessaria: qui la negazione è certa nel suo effetto istituzionale ma incerta nel suo contenuto epistemico, mentre la negazione necessaria è certa nel suo contenuto epistemico ma irrilevante sul piano istituzionale.

La negazione in medicina ha una struttura cognitiva ancora diversa, perché la diagnosi negativa — l'assenza di patologia — non è semplicemente la negazione di una proposizione positiva ma è essa stessa un contenuto informativo che richiede di essere costruito attraverso l'esecuzione di procedure di esclusione sistematica. La diagnosi di esclusione — il processo attraverso cui si arriva alla diagnosi di una condizione rara dopo aver escluso sistematicamente tutte le condizioni più comuni — è un processo di negazione iterata che ha una struttura formale analoga alla semantica di Kripke: si restringe progressivamente l'insieme

dei mondi accessibili escludendo le ipotesi incompatibili con l'evidenza, fino a che rimane un insieme minimale di ipotesi compatibili. Il pensiero diagnostico è, nella sua struttura più profonda, pensiero negativo — pensiero che procede per esclusione sistematica più che per affermazione diretta.

Questa struttura escludente del pensiero diagnostico ha correlati cognitivi precisi che la psicologia del ragionamento medico ha documentato: i medici esperti sono sistematicamente migliori dei novizi nel generare e mantenere ipotesi concorrenti, nel costruire e valutare simulazioni controfattuali dei possibili decorsi, e nel resistere alla chiusura prematura verso un'unica diagnosi prima che l'evidenza la supporti con sufficiente precisione. Queste competenze — il mantenimento del dual-coding diagnostico, la valutazione differenziale della precisione delle ipotesi, la resistenza alla chiusura — sono le stesse competenze cognitive che il framework della negazione proposizionale predice come componenti del sistema maturo di negazione biologica. Il medico esperto è, in questo senso, un esperto della negazione applicata al dominio della fisiologia e della patologia: la sua competenza diagnostica è la sua capacità di negare metodicamente le ipotesi premature, mantenendo aperto lo spazio delle possibilità fino a che l'evidenza converge.

La negazione al livello cellulare — circuiti inibitori, disinibizione e la logica molecolare del non-essere

LXVII. Oltre il livello dei sistemi: la biofisica dell'inibizione come substrato della negazione

Tutto il saggio ha trattato la negazione prevalentemente al livello dei sistemi — reti corticali, aree cerebrali, connessioni tra strutture — con incursioni occasionali nel livello computazionale della codifica predittiva. C'è però un livello di analisi che il saggio originale aveva giustamente identificato come punto di partenza del paradosso — il livello cellulare e molecolare — e che merita ora un'analisi più approfondita, non perché la negazione si riduca alla biofisica sinaptica, ma perché la struttura di quella biofisica pone vincoli precisi su ciò che la negazione può essere a livelli superiori di organizzazione.

La domanda di partenza — come può un sistema fondamentalmente attivazionale rappresentare la negazione come contenuto proposizionale — si trasforma, al livello cellulare, in una domanda più specifica: quali proprietà della microcircuiteria corticale permettono a popolazioni di neuroni di implementare operazioni che, al livello computazionale, corrispondono alla negazione? E la risposta, sorprendentemente, non è univoca: ci sono almeno tre meccanismi cellulari distinti attraverso cui il cervello realizza operazioni funzionalmente equivalenti alla negazione, con proprietà dinamiche e conseguenze computazionali molto diverse.

LXVIII. Gli interneuroni inibitori: una classificazione funzionale

La corteccia cerebrale dei mammiferi contiene approssimativamente tra il 20 e il 25 per cento di interneuroni inibitori — neuroni che rilasciano acido gamma-aminobutirrico (GABA) come neurotrasmettitore principale e che, attraverso i recettori GABA_A ionotropici e GABA_B metabotropici, producono iperpolarizzazione delle cellule bersaglio. Questa cifra è globale; la distribuzione e la funzione degli interneuroni variano enormemente attraverso le diverse aree corticali e gli strati corticali, e questa variazione funzionale è il primo indizio che l'inibizione non è una funzione unitaria ma una famiglia di operazioni computazionalmente distinte.

La classificazione attuale degli interneuroni inibitori corticali — resa possibile dalla convergenza di metodi anatomici, elettrofisiologici, genetici e ottogenetici — distingue tre grandi classi basate sulla loro morfologia assonale e sul profilo di espressione proteica, ciascuna con proprietà dinamiche e bersagli sinaptici caratteristici.

La prima classe è quella delle cellule a canestro (basket cells) e delle cellule a candelabro (chandelier cells), entrambe immunoreattive per la parvalbumina (PV). Le cellule a canestro formano sinapsi inibitorie sui soma e sui segmenti prossimali dei dendriti delle cellule piramidali; le cellule a candelabro formano sinapsi esclusivamente sul segmento iniziale dell'assone — il luogo di generazione del potenziale d'azione — delle cellule piramidali. Queste cellule scaricano a frequenze molto elevate (fast-spiking, tipicamente 100-200 Hz) e seguono con elevata fedeltà i ritmi gamma della rete. La loro funzione computazionale è quella di un'inibizione temporo-sincrona di alta precisione: controllano il timing del firing delle cellule piramidali bersaglio, determinando con esattezza temporale quando una cellula piramidale può o non può scaricare.

Questa proprietà — il controllo del timing piuttosto che della probabilità di firing — è computazionalmente analoga a una negazione temporale puntuale: non dice "questa cellula non deve mai scaricare", ma "questa cellula non deve scaricare adesso". È una negazione contestuale, sensibile al momento e alla fase delle oscillazioni di rete, che implementa la distinzione tra scenari temporali diversi all'interno dello stesso circuito. Le cellule PV sono in questo senso i meccanismi cellulari delle negazioni temporalmente contestualizzate — quelle negazioni che si applicano a uno specifico istante o fase dell'elaborazione piuttosto che all'intero dominio.

La seconda classe è quella delle cellule di Martinotti e delle cellule bipolari e bitufted, immunoreattive per la somatostatina (SST). Queste cellule formano sinapsi inibitorie prevalentemente sui dendriti distali delle cellule piramidali — lontano dal soma, nelle regioni dove arrivano gli input eccitatori cortico-corticali e talamici. La loro funzione computazionale è diversa da quella delle cellule PV: invece di controllare il timing del firing, regolano l'integrazione degli input sui

dendriti, modulando la capacità delle sinapsi eccitatorie distali di influenzare il firing somatico. Questo tipo di inibizione — dendritica, distale, modulatoria — è computazionalmente analoga a una negazione condizionale: non sopprime direttamente l'output della cellula, ma riduce il peso di certi input nella determinazione dell'output.

In termini di codifica predittiva, le cellule SST implementano la modulazione della precisione degli errori di previsione prima che raggiungano i meccanismi di generazione dell'output: agiscono come un filtro che riduce selettivamente il peso di certi segnali ascendenti nella determinazione della risposta del sistema. Questo corrisponde a una negazione epistemica parziale — non la negazione di una proposizione completa, ma la riduzione del grado di credenza in certi input senza eliminarli completamente.

La terza classe è quella dei neuroni immunoreattivi per la calretinina e la calbindina, tra cui i neuroni a doppio bouquet (double-bouquet cells), i neuroni a lampioncino (lamp cells), e i neuroni di Cajal-Retzius nello strato I. Questa classe è la più eterogenea morfologicamente e funzionalmente, e comprende neuroni che formano sinapsi inibitorie su altri interneuroni piuttosto che sulle cellule piramidali — i cosiddetti interneuroni inibitori degli interneuroni, o disinibitori.

LXIX. La disinibizione come doppia negazione cellulare

Il fenomeno della disinibizione — l'inibizione di un interneurone inibitore, che produce come effetto netto un aumento dell'eccitabilità delle cellule piramidali che quello interneurone inibiva — è il meccanismo cellulare della doppia negazione: $\neg(\neg P) \rightarrow P$. Questa operazione, che nella logica classica è una tautologia (la doppia negazione equivale all'affermazione originale), ha al livello biologico una struttura dinamica più ricca e computazionalmente interessante.

Il circuito di disinibizione più documentato nella corteccia cerebrale è quello che coinvolge le cellule VIP (vasoactive intestinal peptide-immunoreactive interneurons) — una sottoclasse della terza classe di interneuroni — come disinibitori delle cellule PV e SST. Quando le cellule VIP si attivano, inibiscono le cellule PV e SST; queste, essendo inibite, riducono la loro influenza inibitoria sulle cellule piramidali; le cellule piramidali, di conseguenza, aumentano la loro eccitabilità. L'effetto netto è un aumento dell'attività delle cellule piramidali prodotto non da input eccitatori diretti, ma dalla rimozione di un'inibizione tonica — la disinibizione appunto.

Questo circuito $VIP \rightarrow PV/SST \rightarrow$ piramidali ha una funzione computazionale precisa che riceve conferma da numerosi studi optogenetici in roditori e primates: segnala l'arrivo di input top-down di alta salienza — input provenienti da strati corticali superiori o da aree corticali di associazione — che richiedono di "aprire" la finestra di elaborazione delle cellule piramidali per permettere l'integrazione di input concorrenti. Le cellule VIP sono, in questo senso, i meccanismi cellulari della "gate apertura" — il segnale che dice al circuito locale

"ora devi ascoltare gli input che stai ricevendo", rimuovendo l'inibizione che normalmente smorza la risposta.

Per la negazione, il circuito di disinibizione ha un'implicazione diretta: permette al sistema di passare da uno stato di inibizione tonica — in cui una rappresentazione è soppressa — a uno stato di attività attraverso un meccanismo che non è l'affermazione diretta della rappresentazione, ma la rimozione della sua negazione. Questo corrisponde esattamente alla struttura formale della doppia negazione classica, ma con una differenza dinamica fondamentale: la rimozione dell'inibizione attraverso disinibizione non produce immediatamente il massimo dell'attivazione, ma apre uno spazio di eccitabilità aumentata in cui l'input eccitatorio può ora produrre firing che prima era soppresso. La disinibizione è una potenzializzazione della possibilità di firing, non il firing stesso — una negazione della negazione che non equivale a un'affermazione completa, ma a un'apertura verso l'affermazione.

Questa distinzione — tra la doppia negazione classica (equivalenza logica con l'affermazione) e la doppia negazione biologica (apertura verso l'affermazione senza equivalenza con essa) — corrisponde alla differenza tra la logica classica, in cui $\neg\neg P = P$, e la logica intuizionista, in cui $\neg\neg P$ implica P ma non è equivalente a P . Il cervello è, in questo senso preciso e cellulare, un sistema intuizionista: la sua doppia negazione non è la stessa cosa dell'affermazione, ma l'apertura dello spazio in cui l'affermazione diventa possibile.

LXX. La Long-Term Potentiation, la Long-Term Depression e la negazione sinaptica permanente

I due meccanismi fondamentali della plasticità sinaptica — la Long-Term Potentiation (LTP) e la Long-Term Depression (LTD) — corrispondono, nella terminologia proposta nella sezione IV del saggio originale seguendo la logica lineare di Girard, rispettivamente alla rafforzamento e all'indebolimento permanente di una connessione sinaptica. Questo è il livello al quale la negazione biologica lascia tracce strutturali durature: non la soppressione momentanea dell'attività attraverso inibizione reversibile, ma la modifica permanente dell'architettura delle connessioni che determina quali informazioni il circuito può e non può elaborare.

La LTD — la riduzione duratura dell'efficacia sinaptica prodotta da pattern di attività specifici, tra cui la stimolazione del postsinaptico in assenza di attività presinaptica o la stimolazione in sequenza inversa rispetto alla regola di Hebb — è il substrato molecolare della negazione irreversibile nel senso della logica lineare: la consumazione definitiva di una risorsa sinaptica, la riduzione permanente del peso con cui una connessione contribuisce all'attività del circuito. Quando il cervello impara che P è falso — quando la credenza in P viene durevolmente negata — il correlato sinaptico è l'indebolimento LTD delle

connessioni che codificavano P come previsione del sistema, e il rafforzamento LTP delle connessioni che codificano $\neg P$.

La molecola di segnalazione più studiata in connessione con la LTD è il recettore NMDA (N-metil-D-aspartato) nella sua condizione di attivazione subthreshold — una situazione in cui il recettore riceve glutammato ma non è sufficientemente depolarizzato da permettere l'ingresso di calcio, producendo invece l'attivazione di cascate di segnalazione che portano all'internalizzazione dei recettori AMPA e all'indebolimento della sinapsi. Questa condizione — il recettore che riceve il segnale ma non può rispondere perché le condizioni di depolarizzazione non sono soddisfatte — è una forma molecolare di frustrazione epistemica: il segnale previsivo arriva (la previsione P viene attivata) ma non trova conferma nell'attività postsinaptica (l'input non corrisponde alla previsione), e questa discrepanza produce non un errore di previsione ascendente bensì una modifica strutturale della sinapsi che riduce il peso della previsione errata. È la forma molecolare della negazione come apprendimento: la correzione permanente di una previsione falsa attraverso la modifica fisica della connessione che la codificava.

La distinzione tra la LTD all'ippocampo e la LTD alla corteccia cerebrale è particolarmente rilevante per la teoria della negazione. La LTD ippocampale ha proprietà di specificità di sequenza — dipende dalla tempistica precisa dell'attivazione presinaptica e postsinaptica in modo Spike Timing-Dependent (STDP) — che la rendono adatta a codificare negazioni episodiche: la correzione di previsioni specifiche associate a eventi particolari nel tempo e nello spazio. La LTD corticale ha proprietà di generalizzazione maggiore — può essere indotta da pattern di attività ripetuta a bassa frequenza senza la precisa specificità temporale dell'STDP — che la rendono adatta a codificare negazioni concettuali: la riduzione del peso di previsioni generalizzate che si sono rivelate sistematicamente false.

Questa distinzione ippocampo/corteccia nella plasticità LTD corrisponde alla distinzione proposta nel saggio tra negazione episodica (che nega un evento specifico nel contesto della memoria episodica) e negazione semantica (che nega una proprietà generalizzata di un concetto): due tipi di negazione che, a livello computazionale e sistemico, appaiono come operazioni dello stesso tipo, ma che a livello molecolare e sinaptico sono implementate attraverso meccanismi parzialmente distinti con proprietà temporali e di generalizzazione diverse.

LXXI. Le oscillazioni di rete e la struttura temporale fine della negazione

Le oscillazioni neurali — i ritmi periodici dell'attività collettiva delle reti corticali — non sono semplici epifenomeni dell'elaborazione neurale ma meccanismi attivi di coordinazione temporale che determinano quali sinapsi possono efficacemente comunicare e in quale finestra temporale. La connessione tra

oscillazioni e negazione è dunque non metaforica ma strutturale: la struttura temporale delle oscillazioni determina la struttura temporale della negazione.

Le oscillazioni gamma (30–80 Hz, con un ciclo di circa 15–30 millisecondi) sono generate dal circuito PING (Pyramidal-Interneuron Network Gamma): le cellule piramidali eccitano le cellule PV, che a loro volta inibiscono sincronamente le piramidali producendo una soppressione temporanea del firing; questa soppressione permette alla depolarizzazione accumulata di raggiungere di nuovo la soglia, producendo un nuovo burst, e così via in un ciclo sostenuto. Ogni ciclo gamma ha dunque una struttura che alterna un periodo di eccitabilità aumentata (la finestra di firing) e un periodo di inibizione (la finestra di soppressione): la struttura temporale più elementare di alternanza presenza-assenza che un sistema oscillatorio possa implementare.

In termini di negazione, ogni ciclo gamma implementa una struttura di affermazione-negazione temporalmente serializzata: nella finestra di eccitabilità, le rappresentazioni attive possono scaricare e contribuire all'output del circuito; nella finestra inibitoria, le stesse rappresentazioni sono soppresse e non contribuiscono. Questa alternanza temporale permette al circuito di implementare la distinzione tra contenuto attivo e contenuto soppresso senza richiedere la coesistenza simultanea di attivazione e inibizione sulla stessa sinapsi — che sarebbe biofisicamente impossibile — ma distribuendo le due fasi nel tempo.

Le oscillazioni theta (4–8 Hz, con un ciclo di 125–250 millisecondi) hanno una struttura diversa e una relazione diversa con la negazione. Come si è discusso in relazione all'ippocampo, le oscillazioni theta organizzano la codifica degli eventi episodici e delle sequenze spaziali in "frame" temporali distinti: diversi eventi che si succedono nella memoria episodica sono codificati in fasi diverse del ciclo theta, permettendo al sistema di mantenere distinte le rappresentazioni di eventi diversi senza confonderli. In termini di negazione, la struttura theta permette al sistema di mantenere contemporaneamente la rappresentazione dell'evento reale (fase theta alta) e la simulazione controfattuale dell'evento non-accaduto (fase theta bassa o opposta), separandole temporalmente nel ciclo theta piuttosto che spazialmente in diverse popolazioni di neuroni. Questa separazione temporale delle fasi theta è il correlato cellulare della struttura husserliana del vuoto intenzionale: la presenza dell'assente in una fase temporale distinta da quella in cui il presente è presente.

L'accoppiamento theta-gamma — il fenomeno per cui i burst gamma si annidano preferenzialmente in fasi specifiche del ciclo theta — fornisce il meccanismo di coordinazione tra i due livelli: la struttura theta organizza la sequenza degli eventi e delle loro simulazioni controfattuali, mentre la struttura gamma all'interno di ogni fase theta determina quale tra le rappresentazioni candidate ottiene la finestra di firing. Questo meccanismo gerarchico di accoppiamento frequenziale è il substrato oscillatorio della struttura gerarchica della negazione proposta nel framework computazionale: la negazione opera a

livelli diversi della gerarchia predittiva, e questi livelli corrispondono a scale temporali diverse nelle oscillazioni di rete.

Biolinguistica, sintassi e la negazione come testa funzionale

LXXII. La sintassi cartografica della negazione: la NegP nella teoria dei sintagmi funzionali

La grammatica generativa — nella sua versione del Minimalismo chomskiano e nella sua elaborazione cartografica da parte di Luigi Rizzi, Guglielmo Cinque e la scuola italiana — ha sviluppato un'analisi della struttura sintattica delle frasi che è di interesse diretto per la neuroscienza della negazione, anche se le connessioni tra i due campi sono rimaste fino ad oggi sorprendentemente poco sviluppate.

Nell'analisi cartografica di Rizzi, elaborata principalmente a partire da Rizzi (1997) e ulteriormente sviluppata nella ricerca successiva, la frase è articolata in una gerarchia di proiezioni funzionali che codificano diversi tipi di informazione strutturale: la forza illocutoria (ForceP), la topicalizzazione (TopP), la focalizzazione (FocP), la negazione (NegP), e la finitezza verbale (FinP), ciascuna come testa funzionale con il proprio specifier e complemento nella struttura X-bar. La posizione della negazione in questa gerarchia — tra la focalizzazione e la finitezza — non è arbitraria: riflette le sue proprietà di interfaccia con la struttura informativa della frase (come si è discusso nella sezione sulla portata e la negazione focalizzata) e con la morfologia verbale (che in molte lingue porta morfema di accordo negativo).

La testa Neg — il nodo sintattico che ospita il morfema negativo — ha proprietà strutturali precise: è una testa funzionale con tratti [NEG] che si propagano attraverso la struttura sintattica attraverso il meccanismo dell'Agree di Chomsky (2000), producendo la concordanza negativa nelle lingue che la ammettono. In italiano, il morfema "non" occupa la testa Neg e i termini a polarità negativa ("nessuno", "mai", "niente") nelle posizioni di argomento o aggiunto sono i termini che entrano in relazione di Agree con la testa Neg, ricevendo da essa il tratto [NEG] che li licenzia come termini negativi piuttosto che come quantificatori positivi.

Questo tratto [NEG] che si propaga attraverso la struttura sintattica è la controparte formale della "negazione distribuita" discussa nella sezione L a proposito della concordanza negativa crosslinguistica: la negazione non è un punto nella struttura sintattica ma un tratto che pervade la struttura, esattamente come la negazione neurale non è un singolo segnale inibitorio ma una modulazione distribuita del tono del processing.

LXXIII. L'interfaccia sintassi-semantica della negazione: la portata nelle derivazioni minimaliste

Il problema della portata della negazione — già affrontato nella sezione II del saggio originale come una delle difficoltà computazionali fondamentali — riceve nell'analisi minimalista un trattamento formale attraverso il concetto di scope assegnato dalla posizione sintattica al livello della Forma Logica (LF). Nel minimalismo, la LF è il livello di rappresentazione sintattica a cui le operazioni semantiche si applicano: la portata relativa dei quantificatori e degli operatori logici è determinata dalle relazioni strutturali — di c-command e scope taking — che i sintagmi hanno al livello LF, prodotto dal movimento sintattico (covert movement) che porta certi costituenti in posizioni di portata superiore.

Per la negazione, questo significa che la portata della negazione rispetto agli altri operatori della frase è determinata dalla posizione che la testa Neg e i costituenti in relazione con essa occupano al livello LF. In italiano, la portata della negazione rispetto ai quantificatori universali e esistenziali produce ambiguità interpretative del tipo già discusso — "non tutti i filosofi sono platonici" può avere la negazione con portata ampia sul quantificatore universale ($\neg \forall x Fx$ = qualcuno non è platonico) o portata stretta sotto il quantificatore ($\forall x \neg Fx$ = nessuno è platonico) — e la disambiguazione richiede sia informazioni prosodiche sia inferenze pragmatiche sul contesto.

La previsione neurolinguistica generata da questa analisi è che il processo di determinazione della portata — la risoluzione dell'ambiguità di scope — debba avere un correlato ERP specifico distinto sia dalla N400 (anomalia semantica) sia dalla P600 (violazione sintattica): una componente associata alla computazione dell'interfaccia sintassi-semantica che determina quale delle interpretazioni possibili è selezionata nel contesto. Questa componente — se esiste come previsto — dovrebbe avere una latenza intermedia tra la N400 e la P600 (circa 500 ms), una distribuzione parietale che riflette l'integrazione tra informazione sintattica e semantica, e una sensibilità specifica al contesto pragmatico che disambigua la portata.

LXXIV. La biolinguistica della negazione: il Merge, la ricorsività e l'origine evolutivista

La biolinguistica — il programma di ricerca che cerca di individuare le basi biologiche del linguaggio umano — ha in Noam Chomsky il suo interlocutore teorico più influente, con la proposta che il nucleo del linguaggio sia un'operazione computazionale minimale — il Merge, l'operazione che combina due elementi linguistici in un'unità strutturata — dotata della proprietà di ricorsività: il Merge può prendere come input il prodotto di un'applicazione precedente del Merge, producendo strutture arbitrariamente annidate.

La negazione, in questa prospettiva, è un prodotto del Merge: il Merge della testa Neg con il suo complemento produce la struttura NegP, e questa struttura può a sua volta essere presa come input dal Merge con altre proiezioni funzionali, producendo strutture di negazione annidata di qualsiasi profondità. La ricorsività del Merge è ciò che permette la negazione della negazione, la negazione di proposizioni che contengono già una negazione, e le strutture di negazione annidata di cui si è parlato nella sezione XXXI.

La domanda biolinguistica fondamentale — quando e come si è evoluta questa capacità — non ha ancora risposta definitiva, ma le evidenze convergono su alcuni vincoli importanti. Il Merge nella sua versione recursiva sembra essere una caratteristica specificatamente umana: i tentativi di insegnare strutture sintatticamente recursive a primati non umani non hanno prodotto risultati convincenti, e la componente di ricorsività sembra essere l'elemento che distingue il linguaggio umano dai sistemi di comunicazione degli altri primati.

Per la negazione, questo implica che la negazione proposizionale ricorsiva — la negazione che può operare su strutture che contengono già negazioni — sia una prerogativa specificatamente umana legata all'emergenza del Merge ricorsivo. Ma il timing evolutivo di questa emergenza è incerto: le evidenze genetiche (in particolare gli studi sul gene FOXP2, associato ai circuiti del linguaggio e della motricità fine orofacciale, e le sue varianti nelle popolazioni umane arcaiche inclusi i Neanderthal) suggeriscono che le basi genetiche del linguaggio siano più antiche del previsto, ma non permettono di determinare se la ricorsività — e con essa la negazione proposizionale — fosse presente nei Neanderthal o sia un'innovazione della linea Homo sapiens.

LXXV. L'interfaccia sintassi-prosodia e la negazione: struttura metrica e curve intonative

La connessione tra la struttura sintattica della negazione e la sua realizzazione prosodica non è arbitraria: è regolata dall'interfaccia sintassi-prosodia, il sistema di mappatura tra struttura sintattica e struttura metrica-intonativa che determina come la struttura sintattica si traduce in pattern di tono, ritmo e accento nel parlato prodotto e percepito.

Per la negazione, questa interfaccia produce previsioni specifiche riguardo ai pattern prosodici associati a diversi tipi di negazione. La negazione a portata ampia — la negazione di un'intera proposizione — è tipicamente associata a un pattern di tono discendente sull'intera frase, con il punto di massima discesa sull'elemento lessicale portante il focus informativo. La negazione focalizzata — la negazione di un singolo costituente con portata stretta — è associata a un picco tonico (accent) sul costituente negato e a un tono piatto o discendente sul resto della frase. Questa correlazione tra struttura intonativa e portata della negazione è documentata attraverso le lingue — con variazioni parametriche nel tono specifico (discendente in lingue a tono basso, ascendente in alcune lingue a

tono alto) — e suggerisce che l'interfaccia sintassi-prosodia mappa sistematicamente le relazioni di portata sintattica sulle strutture dell'intonazione.

Neurologicamente, la produzione di questi pattern prosodici coinvolge l'area supplementare motoria (SMA) e la corteccia motoria primaria per il controllo della muscolatura laringea, con la pianificazione prosodica che dipende dalla corteccia prefrontale sinistra per la struttura ritmica e dalla corteccia frontale destra per la struttura dell'intonazione. La comprensione dei pattern prosodici della negazione coinvolge l'emisfero destro in modo prominente — le aree temporali destre sono specializzate per l'elaborazione della variazione di F0 (frequenza fondamentale) che porta il contorno intonativo — con integrazione nell'emisfero sinistro per l'interfaccia con la struttura semantica.

La previsione specifica è che i pattern prosodici inappropriati per il tipo di negazione — una negazione a portata ampia prodotta con il pattern prosodico della negazione focalizzata, o viceversa — producano difficoltà di comprensione e anomalie ERP che riflettono il conflitto tra le informazioni sintattiche e prosodiche sull'estensione della portata. Questa previsione, verificabile attraverso la presentazione di frasi negative con manipolazione ortogonale del tipo sintattico di negazione e del pattern prosodico, permetterebbe di testare direttamente l'ipotesi che l'interfaccia sintassi-prosodia sia un componente necessario della comprensione della negazione proposizionale — e non soltanto della sua produzione.

LXXVI. La negazione nell'acquisizione del linguaggio: evidenze dalla linguistica acquisizionale

Il terzo stadio ontogenetico della negazione — la negazione proposizionale vera — è stato analizzato nel saggio originale principalmente dal punto di vista delle capacità cognitive che richiede. Ma l'analisi linguistica di come i bambini acquisiscono la struttura sintattica della negazione offre una prospettiva complementare che connette la traiettoria di sviluppo con i meccanismi della grammatica generativa.

I bambini che acquisiscono l'italiano come prima lingua mostrano una traiettoria di acquisizione della negazione sintattica che procede attraverso stadi ben documentati, e questa traiettoria non è arbitraria ma riflette la struttura della grammatica adulta target e le proprietà del processore linguistico in via di maturazione. Il primo stadio — che coincide approssimativamente con la negazione di rifiuto della sezione X — è caratterizzato dall'uso del morfema negativo in posizione esterna alla frase, come operatore che si applica all'intera struttura: "No mangio" anziché "Non mangio", con la negazione che precede l'intera proposizione senza essere integrata nella struttura verbale. Questa strategia — documentata cross linguisticamente nei bambini che acquisiscono lingue con morfologia negativa diversa — è interpretata come l'uso di un

operatore proposizionale esterno alla struttura prima che il bambino abbia acquisito la posizione interna della testa Neg.

Il secondo stadio — che coincide con l'emergenza della negazione di non-esistenza — vede la negazione spostarsi verso l'interno della struttura sintattica: i bambini producono "Non c'è" con la negazione integrata nel verbo copulativo, riflettendo l'acquisizione della posizione sintattica della testa Neg all'interno della frase nucleare. Il terzo stadio — che coincide con la negazione proposizionale vera — è caratterizzato dall'acquisizione delle forme avanzate di negazione: concordanza negativa corretta ("Non ho visto nessuno"), negazione dei termini di polarità positiva nei contesti appropriati, e gestione della portata della negazione in frasi complesse.

La traiettoria acquisizionale della negazione ha, in questa lettura, una doppia natura: è sia la traccia dello sviluppo delle capacità cognitive sottostanti (il cui profilo neurobiologico è stato analizzato nelle sezioni X e XIV) sia la traccia dell'acquisizione della struttura sintattica della negazione nella lingua specifica che si sta apprendendo. La convergenza di questi due processi — maturazione cognitiva e acquisizione linguistica — nella stessa finestra temporale (18–48 mesi) non è casuale: entrambi dipendono dalla maturazione del circuito prefrontale-ippocampale, e il linguaggio — in particolare la struttura sintattica della negazione come testa funzionale — fornisce l'impalcatura esterna che supporta e organizza la maturazione delle capacità cognitive interne.

Questa prospettiva suggerisce un'ipotesi più forte: che l'acquisizione della struttura sintattica della negazione non si limiti a riflettere le capacità cognitive già mature, ma le plasmi — le sostenga e le organizzi durante il loro sviluppo — nel senso vygotskiano della zona di sviluppo prossimale. Il linguaggio, in quanto sistema di negazione simbolica esternamente disponibile, offre al bambino strutture per operare sulla propria cognizione che precedono la piena maturazione dei circuiti neurali che le implementano internamente. L'acquisizione del "non" sintattico non aspetta la maturazione completa del circuito reality-monitoring: la precede, e contribuisce ad essa fornendo un mezzo espressivo e interattivo che incentiva il sistema cognitivo a sviluppare le strutture interne necessarie per quel mezzo.

Questa ipotesi — che il linguaggio favorisca lo sviluppo delle capacità cognitive di negazione anziché limitarsi a esprimerle — è la forma più forte della tesi del determinismo linguistico applicata alla negazione, e offre una prospettiva radicalmente diversa sul rapporto tra linguaggio e cognizione: non il linguaggio come veste della cognizione, non la cognizione come substrato del linguaggio, ma il linguaggio e la cognizione come sistemi in sviluppo mutuo che si sostengono a vicenda attraverso la traiettoria ontogenetica, producendo come risultato una mente che non era interamente prefigurata né nel sistema linguistico né nel sistema cognitivo presi separatamente.

Tre territori ancora inesplorati — attenzione, sonno e implicazioni cliniche

LXXVII. L'attenzione come negazione biologica primordiale

C'è un fenomeno che il saggio ha trattato come sfondo implicito senza mai portarlo in primo piano, benché sia probabilmente la forma più elementare e pervasiva di negazione biologica: l'attenzione selettiva. Quando un sistema nervoso orienta le proprie risorse elaborative verso uno stimolo specifico, lo fa necessariamente a scapito degli altri — la selezione attentiva non è soltanto l'amplificazione del segnale selezionato ma la soppressione degli stimoli concorrenti, sia attraverso inibizione laterale nelle corteccie sensoriali sia attraverso modulazione top-down dai circuiti prefrontali e parietali. L'attenzione è, nella sua struttura computazionale profonda, una negazione: la negazione della rilevanza computazionale di tutto ciò che non è l'oggetto selezionato.

Questa connessione non è metaforica. La teoria del gating sensory di Francis Crick — la proposta che il talamo funzioni come un filtro attivo che nega l'accesso alla corteccia a tutti gli input eccetto quello selezionato dall'attenzione — è una delle formulazioni più esplicite della negazione come meccanismo biologico fondamentale. Il nucleo reticolare del talamo, uno strato di neuroni GABAergici che avvolge il talamo dorsale e riceve collaterali sia dalle proiezioni talamocorticali sia da quelle corticotalamiche, funziona esattamente come un sistema di negazione selettiva: quando l'attenzione si orienta verso un dominio sensoriale specifico, le proiezioni corticali verso il nucleo reticolare sopprimono l'attività dei neuroni relé nelle aree talamiche non selezionate, negando loro l'accesso alla corteccia. Il campo visivo non attenzionato, il canale uditivo non selezionato, il canale somatosensoriale non focalizzato — tutto ciò che non è l'oggetto dell'attenzione viene biologicamente negato prima di raggiungere le corteccie di elaborazione superiore.

Questa negazione attenzionale ha una struttura formale precisa nel framework della codifica predittiva: corrisponde alla riduzione della precisione (precision weighting) degli errori di previsione nei canali non attenzionati. L'attenzione selettiva non sopprime completamente i segnali non selezionati — li de-precisa: riduce il peso con cui i loro errori di previsione influenzano l'aggiornamento del modello generativo. Questo corrisponde esattamente alla negazione epistemica parziale discussa nella sezione LII: non la negazione completa di P, ma la riduzione della sua precisione al di sotto della soglia rilevante per l'aggiornamento del modello.

La connessione tra attenzione e negazione ha implicazioni per la comprensione di alcuni disturbi clinici che combinano anomalie attentive e anomalie della negazione. Nel Disturbo da Deficit di Attenzione e Iperattività (ADHD), le difficoltà di inibizione dell'impulso — la difficoltà a negare comportamentalmente risposte inappropriate — coesistono con le difficoltà di

selezione attentiva — la difficoltà a negare l'elaborazione degli stimoli distraenti. Questa coesistenza non è casuale: riflette la condivisione di substrati neurobiologici (la corteccia prefrontale dorsolaterale e i circuiti dopaminergici mesostriatali) tra il sistema di controllo attentivo e il sistema di inibizione comportamentale. La deficienza dopaminergica nel circuito prefrontale-striatale che caratterizza l'ADHD produce simultaneamente una riduzione della capacità di selezionare attentivamente gli stimoli rilevanti (negare la distrazione) e una riduzione della capacità di inibire le risposte impulsive (negare l'azione). L'ADHD è, nella sua struttura neurobiologica più profonda, un disturbo della negazione biologica — non della negazione proposizionale ma della negazione percettiva e comportamentale che ne costituisce il substrato più elementare.

La meditazione mindfulness — la pratica di portare l'attenzione al momento presente senza giudizio — è, da questa prospettiva, un addestramento sistematico del sistema di negazione attenzionale: l'addestramento a negare la rilevanza dei pensieri che distolgono l'attenzione dal momento presente senza negarli proposizionalmente (senza affermare che i pensieri siano falsi) ma semplicemente senza attribuire loro peso computazionale (riducendone la precisione nel modello generativo). Questa distinzione — tra negare il peso attenzionale di un pensiero e negare il suo contenuto proposizionale — è la struttura cognitiva fondamentale della pratica contemplativa, e corrisponde formalmente alla distinzione tra negazione percettiva (riduzione della precisione dell'input) e negazione proposizionale (modificazione della struttura del modello generativo).

LXXVIII. Il sonno come laboratorio della negazione offline

Il saggio ha trattato la negazione come processo che si svolge durante la veglia — durante la percezione, il ragionamento, la comunicazione. Ma c'è un quarto di ogni giornata in cui il cervello è disconnesso dall'input sensoriale esterno e impegnato in processi che hanno una rilevanza diretta per la teoria della negazione: il sonno, e in particolare le fasi di sonno a onde lente (SWS, Slow-Wave Sleep) e di sonno REM (Rapid Eye Movement).

Durante il SWS, il cervello esegue il processo di consolidamento della memoria: le tracce episodiche registrate durante la veglia nell'ippocampo vengono replicate (reactivated) e progressivamente trasferite verso reti corticali distribuite, dove vengono integrate con la conoscenza semantica preesistente. Questo processo — documentato attraverso la registrazione diretta dell'attività ippocampale e corticale durante il sonno in roditori e, con metodi non invasivi, in esseri umani — non è una semplice copia delle tracce mnestiche esistenti: è un processo di selezione e trasformazione che mantiene certi elementi dell'esperienza e ne sopprime altri.

La selettività del consolidamento mnemonico è una forma di negazione offline: durante il sonno, il sistema decide implicitamente quali componenti dell'esperienza della giornata trascorsa meritano di essere rafforzate nella

memoria a lungo termine e quali devono essere lasciate decadere. Le tracce che vengono consolidate acquisiscono un peso maggiore nel modello generativo; quelle che non vengono consolidate svaniscono — vengono negate nel senso della riduzione permanente del loro peso nel modello. Questa negazione notturna è silenziosa, implicita, priva di contenuto proposizionale consapevole: il cervello addormentato non afferma "questo non è importante" su ciascuna traccia che lascia decadere, ma implementa una selezione la cui struttura formale è quella di una negazione — la riduzione della precisione di certi prior a favore di altri.

Il ruolo del sonno REM nella elaborazione delle esperienze emotivamente cariche aggiunge una dimensione ulteriore. Matthew Walker e collaboratori hanno mostrato che il sonno REM — caratterizzato da basse concentrazioni di noradrenalina, uno dei principali neuromodulatori associati all'arousal e alla risposta di allarme — produce una forma di "estrazione" del contenuto informativo delle esperienze cariche emotivamente separandolo dalla loro componente affettiva acuta. Questo processo — che Walker chiama REM sleep emotional memory processing — corrisponde alla trasformazione di una memoria episodica emotivamente intensa in una memoria semantica emotivamente più neutra: la stessa informazione viene mantenuta, ma la sua precisione affettiva (il peso della risposta emotiva associata) viene ridotta.

Per la negazione, questo processo ha un'implicazione precisa: le negazioni prodotte durante la veglia in risposta a esperienze emotivamente intense — la negazione di un lutto, il rifiuto di una diagnosi, la smentita di un'accusa — vengono elaborate durante il sonno REM in modo da ridurre la loro carica affettiva pur mantenendo il loro contenuto informativo. Il sonno è il mediatore biologico tra la negazione come risposta emotiva acuta e la negazione come struttura cognitiva stabile integrata nel modello generativo: trasforma la negazione reattiva in negazione contemplativa, il rifiuto in accettazione informata.

La connessione tra disturbi del sonno e disturbi della negazione emotiva ha evidenza clinica ben stabilita: il disturbo da stress post-traumatico (PTSD), caratterizzato da incubi ricorrenti e frammentazione del sonno REM, è anche caratterizzato dall'incapacità di estrarre il contenuto informativo delle esperienze traumatiche separandolo dalla loro carica affettiva — esattamente il processo che il sonno REM dovrebbe eseguire. L'ipotesi di Walker — che gli incubi nel PTSD siano tentativi falliti di eseguire il processing REM, interrotti dall'iperattivazione noradrenergica che caratterizza il disturbo — suggerisce che la terapia del PTSD orientata al sonno (inclusa la Image Rehearsal Therapy, che consiste nel riscrivere consapevolmente i contenuti degli incubi durante la veglia) sia una forma di assistenza al processo di negazione offline: aiutare il cervello a eseguire durante la veglia, attraverso la simulazione controfattuale consapevole, ciò che non riesce ad eseguire automaticamente durante il sonno.

Le oscillazioni dei fusi del sonno (sleep spindles, 12–15 Hz) generate dal talamo durante il SWS hanno un ruolo specifico nel consolidamento selettivo: amplificano selettivamente la riattivazione ippocampale delle tracce mnestiche che hanno ricevuto maggiore attenzione durante la veglia, contribuendo alla selezione di ciò che viene consolidato. Nella terminologia del framework proposto, i sleep spindles sono il meccanismo di amplificazione della precisione durante la consolidazione: aumentano il rapporto segnale-rumore delle tracce selezionate, rafforzando la precisione delle previsioni che quelle tracce supportano e lasciando decadere quelle non selezionate. La negazione notturna ha, in questo senso, un ritmo proprio — il ritmo dei fusi del sonno — che è il ritmo biologico della selezione mnestica e, implicitamente, della negazione offline.

LXXIX. Implicazioni cliniche: una sintesi terapeutica

Il framework teorico sviluppato nel saggio non è soltanto di interesse speculativo: genera implicazioni per la pratica clinica che non si limitano alla conferma di trattamenti già esistenti, ma suggeriscono prospettive terapeutiche genuinamente nuove basate sulla struttura computazionale della negazione.

La prima implicazione riguarda i disturbi d'ansia. L'ansia patologica è, nella sua struttura cognitiva, una forma di negazione malfunzionante: il sistema genera previsioni negative — anticipazioni di eventi avversi — con alta precisione e le mantiene come modelli attivi del futuro nonostante l'assenza di evidenze proporzionate alla loro precisione. Questo corrisponde a una negazione del futuro positivo: il sistema nega, con alta precisione non giustificata, la possibilità di esiti favorevoli, producendo un albero temporale in cui le ramificazioni negative hanno un peso sproporzionato. Il trattamento cognitivo-comportamentale dell'ansia — nella sua componente di ristrutturazione cognitiva — è precisamente un intervento sulla struttura del reality monitoring: aiuta il paziente a valutare criticamente la precisione delle proprie previsioni negative, riducendo la loro precisione verso livelli proporzionati all'evidenza disponibile. Il framework proposto suggerisce che l'efficacia di questo trattamento dipenda dalla qualità della simulazione controfattuale che il paziente riesce a costruire: pazienti con maggiore capacità di costruire scenari alternativi vividi e credibili dovrebbero rispondere meglio alla ristrutturazione cognitiva, perché la loro capacità di mantenere modelli controfattuali con sufficiente precisione competitiva permette loro di effettivamente aggiornare il modello generativo ansioso piuttosto che di semplicemente affermare intellettualmente la possibilità di esiti positivi senza integrarla nel modello.

La seconda implicazione riguarda la depressione. La depressione maggiore è caratterizzata, tra le altre cose, da una struttura specifica di negazione del futuro: l'incapacità di costruire simulazioni di futuri positivi con sufficiente precisione da competere con il modello depressogeno. Questo corrisponde a una forma di

impoverimento del sistema di simulazione ippocampale-prefrontale: la connettività funzionale ippocampo-vmPFC è ridotta nella depressione, e questa riduzione corrisponde esattamente alla riduzione della capacità di costruire simulazioni controfattuali di futuri alternativi. La Behavioral Activation Therapy — che incoraggia i pazienti a impegnarsi in attività che precedentemente producevano rinforzo positivo — è, in questa lettura, un intervento sulla simulazione piuttosto che sulla proposizione: non chiede al paziente di negare le credenze depressive (il che richiederebbe un sistema di simulazione già funzionante), ma di accumulare evidenze comportamentali che il sistema può poi usare per aggiornare il modello generativo. È una negazione bottom-up: il comportamento produce input sensoriali che il sistema non può ignorare, forzando un aggiornamento del modello che la ristrutturazione cognitiva top-down non riesce a produrre in pazienti con sistema di simulazione impoverito.

La terza implicazione riguarda il disturbo ossessivo-compulsivo (OOC). Il sintomo cardine del DOC è la difficoltà di negare — di concludere definitivamente che qualcosa è stato fatto, che una porta è chiusa, che le mani sono pulite. Questa incapacità di produrre una negazione definitiva corrisponde, nel framework proposto, a un malfunzionamento del sistema di reality monitoring: il modello controfattuale ("e se la porta fosse aperta") mantiene una precisione competitiva con il modello attuale ("la porta è chiusa") nonostante le evidenze percettive ripetute che la porta è chiusa. Il compito compulsivo — il ri-controllare — è un tentativo di aumentare la precisione del modello attuale attraverso input sensoriali aggiuntivi; ma il DOC è caratterizzato da un deficit nella capacità di accumulare precisione attraverso l'evidenza ripetuta, il che rende il compito compulsivo inutile dal punto di vista computazionale ma irresistibile dal punto di vista dell'urgenza soggettiva.

La terapia di esposizione e prevenzione della risposta (ERP) — il trattamento basato su evidenza per il DOC — è, in questa lettura, un addestramento del sistema di reality monitoring a tollerare la coesistenza di modelli con precisione simile senza tradurla in compulsione: è l'addestramento della paraconsistenza adattiva discussa nella sezione XIX, la capacità di mantenere sia $p(P)$ che $p(\neg P)$ non trascurabili senza intraprendere azioni per ridurre forzatamente l'incertezza. Il meccanismo terapeutico non è la riduzione della precisione del modello ossessivo attraverso la rassicurazione (che avrebbe l'effetto paradossale di aumentare la salienza del dubbio); è l'aumento della tolleranza alla coesistenza di modelli concorrenti — la costruzione di un sistema di negazione robusto che non ha bisogno di certezza assoluta per funzionare.

La quarta implicazione, la più originale rispetto alla letteratura clinica esistente, riguarda il trattamento della confabulazione. Come si è discusso nella sezione XX, la confabulazione è un collasso del differenziale di precisione tra modello attuale e simulazioni ippocampali: il paziente non può distinguere ciò che è realmente accaduto da ciò che ha simulato. Il framework proposto suggerisce che il trattamento della confabulazione non dovrebbe mirare a sopprimere le

confabulazioni attraverso la correzione diretta — che ha l'effetto paradossale di produrre nuove confabulazioni per riempire i vuoti creati dalla correzione — ma a riabilitare il differenziale di precisione attraverso compiti che esercitano specificamente il sistema di reality monitoring prefrontale. Compiti di discriminazione fonte — in cui il paziente deve distinguere tra eventi realmente accaduti ed eventi presentati come ipotesi — potrebbero, con sufficiente pratica e scaffolding contestuale, riabilitare parzialmente il sistema vmPFC di marcatura della realtà, aumentando il differenziale di precisione tra tracce episodiche reali e simulazioni. Questa proposta terapeutica non ha ancora riscontro empirico diretto, ma è generata dal framework con sufficiente precisione da renderne la verifica clinica programmabile.

Appendice A — Derivazioni formali complete

A.1 La semantica di Kripke per la logica intuizionista: derivazione delle condizioni di verità

Un frame di Kripke intuizionista è una tripla $\mathcal{F} = (W, \leq, V)$ dove W è un insieme non vuoto di stati epistemic (mondi possibili), $\leq$ è una relazione di preordine parziale (riflessiva e transitiva) su W che rappresenta l'ordine di accumulazione dell'informazione, e $V : \text{Prop} \times W \rightarrow \{0,1\}$ è una funzione di valutazione che soddisfa la condizione di persistenza (monotonia): se $V(p, w) = 1$ e $w \leq v$ allora $V(p, v) = 1$.

Le condizioni di verità per i connettivi del calcolo proposizionale intuizionista sono:

$V(p, w) = 1$	sse p è atomicamente vero in w (dato)
$V(\perp, w) = 0$	per ogni w ($\perp$ è sempre falso)
$V(A \wedge B, w) = 1$	sse $V(A, w) = 1$ e $V(B, w) = 1$
$V(A \vee B, w) = 1$	sse $V(A, w) = 1$ o $V(B, w) = 1$
$V(A \rightarrow B, w) = 1$	sse per ogni $v \geq w$, se $V(A, v) = 1$ allora $V(B, v) = 1$
$V(\neg A, w) = 1$	sse per ogni $v \geq w$, $V(A, v) = 0$

La negazione intuizionista $\neg A = A \rightarrow \perp$ si ricava dalla condizione per $\rightarrow$:

$V(\neg A, w) = V(A \rightarrow \perp, w) = 1$
sse per ogni $v \geq w$, se $V(A, v) = 1$ allora $V(\perp, v) = 1$
sse per ogni $v \geq w$, $V(A, v) = 0$ (poiché $V(\perp, v) = 0$ sempre)

Questa condizione mostra che $\neg A$ è vera in w soltanto se A non è dimostrabile in nessun futuro stato epistemico accessibile da w — la negazione forte o robusta discussa nella sezione III. Si noti che la condizione $V(\neg\neg A, w) = 1$ non implica $V(A, w) = 1$ in generale (il terzo escluso non vale), ma soltanto che per ogni $v \geq w$ esiste $u \geq v$ tale che $V(A, u) = 1$ — ossia che A è localmente forzabile (locally forced) anche se non attualmente dimostrata.

Lemma (non-equivalenza della doppia negazione intuizionista): Esistono frame di Kripke e proposizioni A tali che $V(\neg\neg A, w) = 1$ e $V(A, w) = 0$.

Dimostrazione. Considera $W = \{w, v\}$ con $w < v$ ($w \leq v$ ma non $v \leq w$), e $V(A, w) = 0$, $V(A, v) = 1$. Allora $V(\neg A, w) = 0$ poiché esiste $v \geq w$ con $V(A, v) = 1$. Dunque per ogni $u \geq w$, $V(\neg A, u) = ?$ Per $u = w$: $V(\neg A, w) = 0$ come appena mostrato. Quindi $V(\neg\neg A, w) = 1$ poiché esiste un futuro u con $V(\neg A, u) = 0$... attendere: la condizione per $\neg\neg A$ richiede che per ogni $v' \geq w$, $V(\neg A, v') = 0$. Calcoliamo: $V(\neg A, v) = 1$ sse per ogni $u \geq v$, $V(A, u) = 0$. Ma $V(A, v) = 1$ e $v \geq v$ (riflessività), quindi $V(\neg A, v) = 0$. Dunque esistono $v' \geq w$ con $V(\neg A, v') = 0$ (prendendo $v' = v$), ma la condizione richiede che per *tutti* i $v' \geq w$ si abbia $V(\neg A, v') = 0$. Poiché $V(\neg A, w) = 0$ e $V(\neg A, v) = 0$, si ha $V(\neg\neg A, w) = 1$. E $V(A, w) = 0$. $\square$

Questa distinzione formale è il fondamento matematico della tesi neuroscientifica: la disinibizione biologica ($\neg\neg P$ biologico) non equivale all'attivazione diretta (P affermato).

A.2 La negazione nella logica lineare: sistema di prova con cut-elimination

La logica lineare di Girard (1987) è formulata come sistema di sequenti con le seguenti regole principali per la negazione duale $(\cdot)^\perp$:

Assiomi:

$$\frac{}{\vdash A, A^\perp} \text{ (Ax)}$$

Taglio:

$$\frac{\vdash \Gamma, A \quad \vdash A^\perp, \Delta}{\vdash \Gamma, \Delta} \text{ (Cut)}$$

Regole strutturali (assenti per le formule lineari, presenti per le esponenziali): La logica lineare non ha indebolimento (Weakening) né

contrazione (Contraction) per le formule non-esponenziali. Le formule sotto $!A$ (bang) o $?A$ (why-not) ammettono invece indebolimento e contrazione:

$$\frac{\vdash \Gamma \quad \vdash \Gamma, ?A, ?A}{\vdash \Gamma, ?A \quad \vdash \Gamma, ?A} \quad \text{(W)} \quad \text{-----} \quad \text{(C)}$$

Le **dualità di de Morgan lineari** si derivano dalle definizioni:

$$\begin{aligned} (A \otimes B) \perp &= A \perp \wp B \perp && \text{(dualità del prodotto tensoriale)} \\ (A \wp B) \perp &= A \perp \otimes B \perp && \text{(dualità del par)} \\ (A \oplus B) \perp &= A \perp \& B \perp && \text{(dualità della somma)} \\ (A \& B) \perp &= A \perp \oplus B \perp && \text{(dualità del with)} \\ (!A) \perp &= ?A \perp && \text{(dualità degli esponenziali)} \\ (?A) \perp &= !A \perp \end{aligned}$$

La **corrispondenza con la biologia neurale** discussa nella sezione IV si formalizza come segue. Sia $E(t)$ il livello di eccitazione e $I(t)$ il livello di inibizione in un circuito neurale al tempo t . La dinamica è:

$$\begin{aligned} dE/dt &= -E/\tau_E + f_E(w_{EE} \cdot E - w_{EI} \cdot I + \text{Input}) \\ dI/dt &= -I/\tau_I + f_I(w_{IE} \cdot E - w_{II} \cdot I) \end{aligned}$$

dove w_{XY} sono i pesi delle connessioni, f_E e f_I sono funzioni di attivazione (tipicamente sigmoidali), e τ_E , τ_I sono le costanti di tempo. L'equilibrio di questo sistema — il punto fisso (E^*, I^*) — corrisponde alla combinazione duale $E \otimes E \perp$ nel senso della logica lineare: E e $E \perp = I$ si annullano mutuamente per produrre lo stato di riposo del circuito. La negazione lineare è letteralmente l'inibizione che bilancia l'eccitazione.

A.3 Il morfismo di negazione bayesiana: derivazione completa

Sia $p(P) \in [0,1]$ la credenza del sistema in P . Nel framework della codifica predittiva gerarchica, la precisione $\Pi(P) = \sigma^{-2}(P)$ è l'inverso della varianza della distribuzione di credenza in P . La negazione bayesiana produce $p(\neg P) = 1 - p(P)$, con precisione:

$$\Pi(\neg P) = (dp(\neg P)/dp(P))^2 \cdot \Pi(P) = \Pi(P)$$

poiché la trasformazione è lineare con derivata -1 . Tuttavia questa forma semplice non cattura la struttura della negazione proposizionale che dipende dalla qualità della prova di negazione. Si introduce quindi il morfismo arricchito della sezione XVIII:

$$\Pi(\neg P) = f(\Pi(P), \Pi(P \rightarrow \perp), \Pi(M))$$

dove:

- $\Pi(P)$ = precisione della proposizione da negare
- $\Pi(P \rightarrow \perp)$ = precisione della derivazione di assurdità da P (la "prova" della negazione)
- $\Pi(M)$ = precisione del modello generativo M

Derivazione delle condizioni di coerenza. La funzione f deve soddisfare:

(i) *Quando $\Pi(P \rightarrow \perp) \rightarrow \infty$ e $\Pi(M) \rightarrow \infty$:* La negazione è basata su una prova perfetta in un modello perfetto, quindi $\Pi(\neg P) \rightarrow \infty$. Formalmente: $\lim_{\{\Pi_{\text{proof}} \rightarrow \infty, \Pi_M \rightarrow \infty\}} f = \infty$.

(ii) *Quando $\Pi(P \rightarrow \perp) = 0$ (nessuna prova):* Il sistema non ha prova della negazione; si affida al solo prior del modello: $f(\Pi, 0, \Pi_M) = \Pi_M / (1 + \Pi / \Pi_M) = \Pi_M^2 / (\Pi_M + \Pi)$.

(iii) *Quando $\Pi(P) = 0$ (P completamente incerta):* Qualsiasi prova, anche debole, di $\neg P$ è sufficiente: $f(0, \Pi_{\text{proof}}, \Pi_M) = \Pi_{\text{proof}}$.

Una forma chiusa che soddisfa tutte e tre le condizioni è:

$$\Pi(\neg P) = \frac{\Pi_M + \Pi_{\text{proof}}}{1 + \Pi / \Pi_M}$$

Verifica delle condizioni:

(i): quando $\Pi_{\text{proof}} \rightarrow \infty$ e $\Pi_M \rightarrow \infty$, numeratore $\rightarrow \infty$, denominatore $\rightarrow 1 + \Pi / \infty = 1$, quindi $\Pi(\neg P) \rightarrow \infty$. ✓

(ii): $f(\Pi, 0, \Pi_M) = \Pi_M / (1 + \Pi / \Pi_M) = \Pi_M^2 / (\Pi_M + \Pi)$. ✓

(iii): $f(0, \Pi_{\text{proof}}, \Pi_M) = (\Pi_M + \Pi_{\text{proof}}) / (1 + 0) = \Pi_M + \Pi_{\text{proof}} \approx \Pi_{\text{proof}}$ quando $\Pi_{\text{proof}} \gg \Pi_M$. ✓ (con un termine residuo Π_M che rappresenta il prior del modello, biologicamente ragionevole)

Interpretazione. Il morfismo ha la struttura di un aggiornamento bayesiano: il numeratore somma i contributi indipendenti del modello (Π_M) e della prova di negazione (Π_{proof}); il denominatore penalizza la precisione della negazione in proporzione alla precisione di P (più P è precisa, più difficile è negarla — coerentemente con la psicologia del ragionamento).

A.4 La variational free energy e la negazione: espansione formale

Il principio di energia libera di Friston definisce la variational free energy come:

$$\begin{aligned} F &= E_q[\log q(x) - \log p(x, o)] \\ &= E_q[\log q(x)] - E_q[\log p(o|x)] - E_q[\log p(x)] \\ &= -H[q(x)] - E_q[\log p(o|x)] - E_q[\log p(x)] \end{aligned}$$

dove il primo termine è l'entropia negativa (o negentropia) della distribuzione variazionale, il secondo è la log-verosimiglianza attesa, e il terzo è la log-prior attesa. Minimizzare F rispetto a q equivale a massimizzare il lower bound sulla log-evidenza:

$$\begin{aligned} \log p(o) &\geq -F = E_q[\log p(o|x)] + E_q[\log p(x)] - E_q[\log q(x)] \\ &= E_q[\log p(o|x)] - D_{KL}[q(x) \parallel p(x)] \end{aligned}$$

Per la negazione, consideriamo due modelli concorrenti: M (il modello attuale, che prevede $\neg P$) e M' (il modello controfattuale, che prevede P). L'energia libera di ciascun modello è:

$$\begin{aligned} F_M &= E_{q_M}[\log q_M(x)] - E_{q_M}[\log p(o|x, \neg P)] - E_{q_M}[\log p(x|\neg P)] \\ F_{M'} &= E_{q_{M'}}[\log q_{M'}(x)] - E_{q_{M'}}[\log p(o|x, P)] - E_{q_{M'}}[\log p(x|P)] \end{aligned}$$

La struttura della negazione proposizionale nella sezione XII è implementata dal differenziale:

$$\Delta F = F_{M'} - F_M > 0$$

ossia il modello controfattuale M' (che prevede P) ha energia libera maggiore del modello attuale M (che prevede $\neg P$). La differenza ΔF misura la robustezza della negazione: tanto più grande è ΔF , tanto più costoso sarebbe per il sistema abbandonare il modello $\neg P$ in favore di P . La precisione differenziale discussa nella sezione XII è:

$$\Delta \Pi = \Pi_M - \Pi_{M'} = \Sigma_M^{-1} - \Sigma_{M'}^{-1} > 0$$

con Σ_M la covarianza del modello attuale (piccola, alta precisione) e $\Sigma_{M'}$ la covarianza del modello controfattuale (grande, bassa precisione).

A.5 La negazione nella logica temporale ramificata: dualità formali

In CTL* (Computation Tree Logic con path quantifiers), la struttura di un sistema di transizione $M = (S, \rightarrow, L)$ con stati S , relazione di transizione $\rightarrow$ e labeling L definisce le seguenti condizioni di verità per i path quantifiers A (tutti i cammini) ed E (esiste un cammino) combinati con gli operatori temporali G (globally), F (finally), X (next), U (until):

- $M, s \models A \varphi$ sse per ogni cammino π da s , $M, \pi \models \varphi$
- $M, s \models E \varphi$ sse esiste un cammino π da s tale che $M, \pi \models \varphi$
- $M, \pi \models G \varphi$ sse per ogni $i \geq 0$, $M, \pi[i..] \models \varphi$
- $M, \pi \models F \varphi$ sse esiste $i \geq 0$ tale che $M, \pi[i..] \models \varphi$
- $M, \pi \models \varphi U \psi$ sse esiste $i \geq 0$ tale che $M, \pi[i..] \models \psi$
e per ogni $0 \leq j < i$, $M, \pi[j..] \models \varphi$

Le **dualità fondamentali** che corrispondono alla struttura della negazione temporale (sezione XLIII) sono:

- $\neg A G \varphi \equiv E F \neg \varphi$ (non necessariamente sempre: su qualche cammino, prima o poi non)
- $\neg A F \varphi \equiv E G \neg \varphi$ (non necessariamente prima o poi: su qualche cammino, sempre non)
- $\neg E G \varphi \equiv A F \neg \varphi$ (su nessun cammino sempre: su tutti i cammini, prima o poi non)
- $\neg E F \varphi \equiv A G \neg \varphi$ (su nessun cammino prima o poi: su tutti i cammini, sempre non)
- $\neg(\varphi U \psi) \equiv (\neg\psi U (\neg\varphi \wedge \neg\psi)) \vee G \neg\psi$ (negazione del until)

La corrispondenza con la psicologia del ragionamento probabilistico (sezione XLIII) è: la violazione della dualità $\neg A F \varphi \equiv E G \neg \varphi$ corrisponde alla confusione tra "non necessariamente prima o poi P " (= esiste un futuro in cui P non accade mai) e "mai P " (= in tutti i futuri, P non accade mai). Questo è esattamente l'errore di overestimazione della negazione universale documentato negli studi di Kahneman e Tversky sui bias cognitivi nella valutazione delle probabilità.

Appendice B — Paradigmi sperimentali specificati

B.1 Paradigma per la distinzione negazione necessaria / accidentale in ERP

Materiali. 120 coppie di stimoli: 40 frasi con negazione di proprietà necessarie (es. "Un cerchio non ha angoli"), 40 frasi con negazione di proprietà accidentali contestualmente stabilite (es. "Mario non era a Roma ieri"), 40 frasi affermative di controllo. Tutte le frasi sono equilibrate e di frequenza lessicale controllata (normate sul corpus ItTenTen).

Procedura. Presentazione seriale rapida (RSVP) a 300 ms per parola. I soggetti eseguono un compito di verità/falsità con risposta manuale (tasto sinistro = vero, tasto destro = falso). EEG a 64 canali, frequenza di campionamento 1024 Hz.

Variabili dipendenti critiche. (1) Ampiezza e latenza della N400 alla parola critica (la parola che determina il tipo di negazione) per i tre tipi di stimolo. (2) Distribuzione topografica della LFN (300–600 ms) con attenzione alla componente frontale vs. centrale. (3) Latenza della risposta comportamentale.

Previsione specifica. La negazione necessaria produrrà una N400 con picco 20–40 ms più tardivo rispetto alla negazione accidentale, distribuzione più anteriore (Fz, F3, F4 > Cz, Pz), e latenza di risposta comportamentale 50–80 ms più breve. Il risultato alternativo falsificante: assenza di differenza nella latenza della N400 ($t < 1.5$ ms, $p > .10$ in analisi cluster-based permutation).

B.2 Paradigma per il dual-coding ippocampale in fMRI

Materiali. Frasi in tre condizioni: (A) negazione accidentale ricca con contenuto controfattuale simulabile (es. "Francesca non è andata a Berlino, avrebbe potuto visitare il Reichstag"); (B) negazione necessaria senza contenuto controfattuale simulabile (es. "Un numero primo non è divisibile per 4, per definizione"); (C) affermazioni di controllo equalizzate per lunghezza e frequenza.

Procedura. fMRI event-related a 3T, TR 2s, 36 slice assiali, matrice 64×64, spessore 3.5mm. I soggetti leggono le frasi e premono un tasto al termine della comprensione. Analisi: ROI-based su ippocampo (segmentato con FreeSurfer), vmPFC, e dlPFC; Dynamic Causal Modelling per la stima della connettività ippocampo-vmPFC.

Previsione specifica. La condizione (A) produrrà attivazione ippocampale bilaterale significativamente maggiore della condizione (B) (differenza attesa: $\beta \geq 0.3$ nell'analisi ROI, FWE-corrected). La connettività funzionale ippocampo-vmPFC sarà maggiore in (A) rispetto a (B) con parametro DCM positivo per la connessione modulatoria (media di gruppo > 0.05 Hz, t-test one-sample $p < .01$). Risultato falsificante: assenza di differenza ippocampale tra (A) e (B), o connettività ippocampo-vmPFC maggiore in (B) che in (A).

B.3 Paradigma per la compatibilità motoria della negazione temporale

Materiali. 80 frasi in quattro condizioni: (A) negazione immediata di azione motoria (es. "Non afferrare la tazza sul tavolo"); (B) negazione temporalmente distante di azione motoria (es. "Quando entrerai nella cucina non afferrare la tazza sul tavolo"); (C) negazione immediata senza contenuto motorio (es. "Non è vero che il cielo è verde"); (D) affermazione motoria di controllo (es. "Afferra la tazza sul tavolo"). Seguita da immagine probe di un'azione (afferrare o spingere) rispetto alla quale i soggetti indicano compatibilità/incompatibilità.

Previsione specifica. L'effetto di compatibilità motoria (RT per azione compatibile vs. incompatibile con il contenuto motorio) sarà presente e di ampiezza simile per le condizioni (A) e (D), mentre sarà significativamente ridotto ($< 50\%$ dell'ampiezza in A) per la condizione (B). Questo effetto di interazione Tipo di negazione $\times$ Distanza temporale è la previsione critica che nessun modello puramente semantico (senza struttura temporale) può generare.

Appendice C — Glossario dei formalismi

Attractor — punto fisso stabile o ciclo limite nel sistema dinamico del modello generativo; una configurazione di stati neurali verso cui il sistema converge. La negazione proposizionale corrisponde alla transizione tra attrattori.

Concordanza negativa — fenomeno crosslinguistico per cui più costituenti di una frase portano marcatura negativa, producendo una sola negazione semantica nonostante la molteplicità formale dei morfemi negativi.

CTL* — Computation Tree Logic, logica temporale su strutture ad albero ramificato che permette di esprimere proprietà dei cammini computazionali; usata nella sezione XLIII per formalizzare la negazione di traiettorie future.

D_{KL} — Divergenza di Kullback-Leibler: $D_{KL}[q \parallel p] = \int q(x) \log(q(x)/p(x)) dx$; misura la distanza informazionale tra due distribuzioni di probabilità. Appare nella variational free energy come misura della distanza tra il modello del sistema e la distribuzione posteriore ideale.

Disinibizione — il meccanismo cellulare per cui l'inibizione di un interneurone inibitore produce un aumento netto dell'eccitabilità delle cellule piramidali bersaglio; corrisponde alla doppia negazione biologica ($\neg\neg P \rightarrow$ maggiore accessibilità di P senza equivalenza con P).

Energy landscape — paesaggio energetico: la distribuzione dell'energia libera nello spazio degli stati del modello generativo; i minimi locali sono gli attrattori, le barriere di potenziale tra minimi determinano la robustezza delle negazioni.

Energia libera variazionale — $F = E_q[\log q - \log p(x, \theta)]$; il functional che i sistemi biologici minimizzano secondo il principio di Friston; equivale a un upper bound sulla surprisal logaritmica.

Frame di Kripke — struttura formale (W, R, V) o $(W, \leq, V)$ usata per la semantica delle logiche modali e intuizioniste; W è l'insieme dei mondi possibili, R o $\leq$ è la relazione di accessibilità, V è la funzione di valutazione.

LFN — Late Frontal Negativity: componente ERP negativa nelle regioni frontali con latenza 300–500 ms, associata specificamente al costo computazionale della negazione proposizionale.

Logica FOUR (Belnap) — sistema logico a quattro valori $\{T, F, B, N\}$ (vero, falso, entrambi, nessuno); la negazione è: $\neg T = F$, $\neg F = T$, $\neg B = B$, $\neg N = N$. Discutito nelle sezioni XIX e LIV in connessione con il catuskoti nagarjuniano.

LTD/LTP — Long-Term Depression / Long-Term Potentiation: riduzioni e aumenti persistenti dell'efficacia sinaptica prodotti da pattern specifici di attività; implementano rispettivamente la negazione sinaptica irreversibile e il rafforzamento della credenza.

Merge — operazione computazionale minimale della grammatica generativa che combina due oggetti sintattici in uno strutturato; la sua ricorsività è il fondamento della negazione proposizionale annidata.

N400 — componente ERP negativa con picco intorno a 400 ms, sensibile alle anomalie semantiche e agli errori di previsione semantici; usata come misuratore dell'errore di previsione semantico nel testo.

NegP — Negation Phrase: proiezione funzionale della struttura sintattica che ospita il morfema negativo come testa funzionale nella cartografia rizzeana della frase.

Paraconsistenza — proprietà di un sistema logico o cognitivo che tollera la coesistenza di contraddizioni locali senza propagarle a tutto il sistema (ex contradictione quodlibet non vale); discussa come proprietà adattiva del cervello nella sezione XIX.

Precision weighting — pesatura della precisione: il meccanismo attraverso cui il sistema della codifica predittiva regola il peso degli errori di previsione in funzione della loro precisione stimata; è il meccanismo neurale dell'attenzione selettiva e del reality monitoring.

Reality monitoring — il processo cognitivo che distingue le rappresentazioni di origine percettiva (reale) da quelle di origine immaginativa (simulata o controfattuale); implementato principalmente dalla vmPFC; discusso nelle sezioni XII e LXXX.

Semantica dei giochi (Hintikka) — semantica in cui la verità di una proposizione è definita dall'esistenza di una strategia vincente per il Verificatore in un gioco semantico contro il Falsificatore; la negazione è lo scambio dei ruoli tra i due giocatori.

Sleep spindles — fusi del sonno: oscillazioni talamiche a 12–15 Hz durante il SWS; coinvolti nel consolidamento selettivo della memoria episodica; proposti nella sezione LXXVIII come substrato del "ritmo della negazione offline".

Śūnyatā — termine sanscrito per vacuità o assenza di natura intrinseca nel pensiero buddhista Madhyamaka; discusso nella sezione LIV come analogo della zona di indeterminatezza nella logica FOUR e nelle strutture paraconsistenti.

Topos — categoria cartesianamente chiusa con un classificatore di sottooggetti Ω ; generalizza la struttura degli insiemi permettendo logiche interne non-booleane; usato nelle sezioni XVI e XVII per la formalizzazione del classificatore di verità neurale continuo.

vmPFC / dlPFC — corteccia prefrontale ventromediale / dorsolaterale: la prima coinvolta nel reality monitoring e nella valutazione delle simulazioni; la seconda nel mantenimento attivo di rappresentazioni multiple in memoria di lavoro; entrambe componenti del circuito neurale della negazione proposizionale matura.

Bibliografia

La bibliografia è organizzata per area tematica, in corrispondenza con le parti del saggio. All'interno di ciascuna sezione l'ordine è alfabetico per autore.

I. Logica formale e fondamenti matematici della negazione

Belnap, N. D. (1977). A useful four-valued logic. In J. M. Dunn & G. Epstein (Eds.), *Modern Uses of Multiple-Valued Logic* (pp. 5–37). Springer.

Brouwer, L. E. J. (1907). *Over de Grondslagen der Wiskunde*. Maas & van Suchtelen. [Trad. it. parziale: Sui fondamenti della matematica.]

da Costa, N. C. A. (1974). On the theory of inconsistent formal systems. *Notre Dame Journal of Formal Logic*, 15(4), 497–510.

Frege, G. (1919). Die Verneinung. *Beiträge zur Philosophie des deutschen Idealismus*, 1, 143–157. [Trad. it.: La negazione. In *Ricerche Logiche*. Guerini, 1988.]

Girard, J.-Y. (1987). Linear logic. *Theoretical Computer Science*, 50(1), 1–102.

Heyting, A. (1930). Die formalen Regeln der intuitionistischen Logik. *Sitzungsberichte der Preussischen Akademie der Wissenschaften*, 42–56.

Hintikka, J. (1968). Language-games for quantifiers. *American Philosophical Quarterly Monograph Series*, 2, 46–72.

Horn, L. R. (1989). *A Natural History of Negation*. University of Chicago Press. [Trad. it.: *Una storia naturale della negazione*. La Nuova Italia, 2000.]

Kripke, S. A. (1963). Semantical considerations on modal logic. *Acta Philosophica Fennica*, 16, 83–94.

Lawvere, F. W., & Schanuel, S. (1997). *Conceptual Mathematics: A First Introduction to Categories*. Cambridge University Press.

Priest, G. (1987). *In Contradiction: A Study of the Transconsistent*. Nijhoff. [2ª ed. ampliata: Oxford University Press, 2006.]

Prior, A. N. (1957). *Time and Modality*. Oxford University Press.

Troelstra, A. S., & van Dalen, D. (1988). *Constructivism in Mathematics: An Introduction* (2 voll.). North-Holland.

II. Fenomenologia e filosofia della negazione

Frege, G. (1919). Op. cit.

Husserl, E. (1900–1901). *Logische Untersuchungen* (2 voll.). Niemeyer. [Trad. it.: *Ricerche logiche*. Il Saggiatore, 2005.]

Husserl, E. (1913). *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie*, Bd. I. Niemeyer. [Trad. it.: *Idee per una fenomenologia pura*. Einaudi, 2002.]

Husserl, E. (1928). *Vorlesungen zur Phänomenologie des inneren Zeitbewusstseins*. Niemeyer. [Trad. it.: *Per la fenomenologia della coscienza interna del tempo*. Franco Angeli, 1998.]

Nagarjuna (ca. II sec. d.C.). *Mulamadhyamakakarika*. [Trad. it.: *I versi del cuore del cammino di mezzo*. Neri Pozza, 2013.]

Priest, G., Garfield, J., & Deguchi, Y. (2018). *The Moon Points Back*. Oxford University Press.

Ricœur, P. (1990). *Soi-même comme un autre*. Seuil. [Trad. it.: *Sé come un altro*. Jaca Book, 1993.]

Sartre, J.-P. (1943). *L'être et le néant*. Gallimard. [Trad. it.: *L'essere e il nulla*. Il Saggiatore, 2002.]

III. Neuroscienze cognitive della negazione: studi fondamentali

Bahlmann, J., Mueller, J. L., Makuuchi, M., & Friederici, A. D. (2011). Perisylvian functional connectivity during processing of sentential negation. *Frontiers in Psychology*, 2, 104.

Coltheart, M., Langdon, R., & McKay, R. (2011). Delusional belief. *Annual Review of Psychology*, 62, 271–298.

Fischler, I., Bloom, P. A., Childers, D. G., Roucos, S. E., & Perry, N. W. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology*, 20(4), 400–409.

Glenberg, A. M., Havas, D., Becker, R., & Rinck, M. (2005). Grounding language in bodily states. In R. Zwaan & D. Pecher (Eds.), *The Grounding of Cognition* (pp. 115–128). Cambridge University Press.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304(5669), 438–441.

Kaup, B., Lüdtke, J., & Zwaan, R. A. (2006). Processing negated sentences with contradictory predicates. *Brain and Language*, 96(2), 149–162.

Nieuwland, M. S., & Kuperberg, G. R. (2008). When the truth is not too hard to handle. *Psychological Science*, 19(12), 1213–1218.

Nieuwland, M. S., & van Berkum, J. J. A. (2006). When peanuts fall in love. *Journal of Cognitive Neuroscience*, 18(7), 1098–1111.

Tettamanti, M., Manenti, R., Della Rosa, P. A., Falini, A., Perani, D., Cappa, S. F., & Moro, A. (2008). Negation in the brain. *NeuroImage*, 43(2), 358–367.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally simulate verbal motion events. *Psychological Science*, 13(2), 168–171.

IV. Codifica predittiva e principio di energia libera

Clark, A. (2016). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press. [Trad. it.: *Surfing sull'incertezza*. Cortina, 2019.]

Friston, K. J. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B*, 360(1456), 815–836.

Friston, K. J. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138.

Friston, K. J., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology-Paris*, 100(1–3), 70–87.

Helmholtz, H. von (1867). *Handbuch der physiologischen Optik*, Bd. 3. Voss. [Sulla percezione come inferenza inconscia.]

Hohwy, J. (2013). *The Predictive Mind*. Oxford University Press.

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex. *Nature Neuroscience*, 2(1), 79–87.

V. Memoria, ippocampo e simulazione controfattuale

Addis, D. R., Wong, A. T., & Schacter, D. L. (2007). Remembering the past and imagining the future. *Neuropsychologia*, 45(7), 1363–1377.

Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in Cognitive Sciences*, 11(2), 49–57.

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network. *Annals of the New York Academy of Sciences*, 1124, 1–38.

Doll, B. B., Duncan, K. D., Simon, D. A., Shohamy, D., & Daw, N. D. (2015). Model-based choices involve prospective neural activity. *Nature Neuroscience*, 18(5), 767–772.

O'Keefe, J., & Nadel, L. (1978). *The Hippocampus as a Cognitive Map*. Oxford University Press.

Pfeiffer, B. E., & Foster, D. J. (2013). Hippocampal place-cell sequences depict future paths to remembered goals. *Nature*, 497(7447), 74–79.

Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory. *Philosophical Transactions of the Royal Society B*, 362(1481), 773–786.

VI. Sviluppo ontogenetico della negazione

Bloom, L. (1970). *Language Development: Form and Function in Emerging Grammars*. MIT Press.

Drozd, K. F. (2001). Children's weak interpretations of universally quantified questions. In M. Bowerman & S. C. Levinson (Eds.), *Language Acquisition and Conceptual Development* (pp. 340–376). Cambridge University Press.

Perner, J., Leekam, S. R., & Wimmer, H. (1987). 'Three-year-olds' difficulty with false belief. *British Journal of Developmental Psychology*, 5(2), 125–137.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs. *Cognition*, 13(1), 103–128.

VII. Logica modale e aspetti metafisici

Chalmers, D. J. (2002). Does conceivability entail possibility? In T. S. Gendler & J. Hawthorne (Eds.), *Conceivability and Possibility* (pp. 145–200). Oxford University Press.

Kripke, S. A. (1980). *Naming and Necessity*. Harvard University Press. [Trad. it.: *Nome e necessità*. Bollati Boringhieri, 1999.]

Lewis, D. K. (1986). *On the Plurality of Worlds*. Blackwell.

Quine, W. V. O. (1953). Reference and modality. In *From a Logical Point of View* (pp. 139–159). Harvard University Press.

VIII. Teoria degli atti linguistici e pragmatica

Austin, J. L. (1962). *How to Do Things with Words*. Oxford University Press. [Trad. it.: *Come fare cose con le parole*. Marietti, 1987.]

Fauconnier, G. (1985). *Mental Spaces*. MIT Press. [Trad. it.: *Spazi mentali*. Anabasi, 1994.]

Fauconnier, G., & Turner, M. (2002). *The Way We Think*. Basic Books.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics, Vol. 3: Speech Acts* (pp. 41–58). Academic Press.

Searle, J. R. (1969). *Speech Acts*. Cambridge University Press. [Trad. it.: *Atti linguistici*. Bollati Boringhieri, 1992.]

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition*. Blackwell. [2^a ed. 1995. Trad. it.: *La pertinenza*. Anabasi, 1993.]

IX. Tipologia crosslinguistica e universali

Dryer, M. S. (1988). Universals of negative position. In M. Hammond, E. Moravcsik, & J. Wirth (Eds.), *Studies in Syntactic Typology* (pp. 93–124). Benjamins.

Greenberg, J. H. (1966). Language universals. In T. A. Sebeok (Ed.), *Current Trends in Linguistics*, Vol. 3. Mouton.

Horn, L. R. (1989). Op. cit.

Payne, J. R. (1985). Negation. In T. Shopen (Ed.), *Language Typology and Syntactic Description*, Vol. 1 (pp. 197–242). Cambridge University Press.

X. Neuropsicologia della negazione: patologia

Dalla Barba, G. (1993). Different patterns of confabulation. *Cortex*, 29(4), 567–581.

Feinberg, T. E., Haber, L. D., & Leeds, N. E. (1990). Verbal asomatognosia. *Neurology*, 40(9), 1391–1394.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, 114(1), 3–28.

Moscovitch, M. (1995). Confabulation. In D. L. Schacter (Ed.), *Memory Distortion* (pp. 226–251). Harvard University Press.

XI. Emozione, soppressione e regolazione emotiva

Gross, J. J. (1998). Antecedent- and response-focused emotion regulation. *Journal of Personality and Social Psychology*, 74(1), 224–237.

Ochsner, K. N., & Gross, J. J. (2005). The cognitive control of emotion. *Trends in Cognitive Sciences*, 9(5), 242–249.

Wegner, D. M., Schneider, D. J., Carter, S. R., & White, T. L. (1987). Paradoxical effects of thought suppression. *Journal of Personality and Social Psychology*, 53(1), 5–13.

XII. Neurobiologia cellulare dell'inibizione

Freund, T. F., & Buzsáki, G. (1996). Interneurons of the hippocampus. *Hippocampus*, 6(4), 347–470.

Harris, K. D., & Mrsic-Flogel, T. D. (2013). Cortical connectivity and sensory coding. *Nature*, 503(7474), 51–58.

Kepecs, A., & Fishell, G. (2014). Interneuron cell types are fit to function. *Nature*, 505(7483), 318–326.

Pi, H.-J., Hangya, B., Kvitsiani, D., Sanders, J. I., Huang, Z. J., & Kepecs, A. (2013). Cortical interneurons that specialize in disinhibitory control. *Nature*, 503(7477), 521–524.

Somogyi, P., & Klausberger, T. (2005). Defined types of cortical interneurone structure space and spike timing in the hippocampus. *Journal of Physiology*, 562(1), 9–26.

XIII. Oscillazioni neurali

Buzsáki, G. (2006). *Rhythms of the Brain*. Oxford University Press.

Buzsáki, G., & Wang, X.-J. (2012). Mechanisms of gamma oscillations. *Annual Review of Neuroscience*, 35, 203–225.

Fries, P. (2005). A mechanism for cognitive dynamics. *Trends in Cognitive Sciences*, 9(10), 474–480.

Lisman, J. E., & Jensen, O. (2013). The theta-gamma neural code. *Neuron*, 77(6), 1002–1016.

Wang, X.-J. (2010). Neurophysiological and computational principles of cortical rhythms in cognition. *Physiological Reviews*, 90(3), 1195–1268.

XIV. Linguistica generativa e biolinguistica

Chomsky, N. (1995). *The Minimalist Program*. MIT Press.

Chomsky, N. (2000). Minimalist inquiries. In R. Martin, D. Michaels, & J. Uriagereka (Eds.), *Step by Step* (pp. 89–155). MIT Press.

Cinque, G. (1999). *Adverbs and Functional Heads*. Oxford University Press.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language. *Science*, 298(5598), 1569–1579.

Rizzi, L. (1997). The fine structure of the left periphery. In L. Haegeman (Ed.), *Elements of Grammar* (pp. 281–337). Kluwer.

Tomasello, M. (2003). *Constructing a Language*. Harvard University Press.

XV. Sonno, consolidamento e memory processing

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience*, 11(2), 114–126.

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*, 437(7063), 1272–1278.

Walker, M. P. (2009). The role of sleep in cognition and emotion. *Annals of the New York Academy of Sciences*, 1156, 168–197.

Walker, M. P., & van der Helm, E. (2009). Overnight therapy? The role of sleep in emotional brain processing. *Psychological Bulletin*, 135(5), 731–748.

XVI. Attenzione, ADHD e mindfulness

Crick, F. (1984). Function of the thalamic reticular complex. *Proceedings of the National Academy of Sciences*, 81(14), 4586–4590.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, 13, 25–42.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18, 193–222.

Tang, Y.-Y., Hölzel, B. K., & Posner, M. I. (2015). The neuroscience of mindfulness meditation. *Nature Reviews Neuroscience*, 16(4), 213–225.

XVII. Filosofia della mente e coscienza

Chalmers, D. J. (1996). *The Conscious Mind*. Oxford University Press. [Trad. it.: *La mente cosciente*. McGraw-Hill, 1999.]

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, 83(4), 435–450.

Searle, J. R. (1992). *The Rediscovery of the Mind*. MIT Press. [Trad. it.: *La riscoperta della mente*. Bollati Boringhieri, 1994.]

XVIII. Teoria del topos e logica categoriale

Goldblatt, R. (1984). *Topoi: The Categorical Analysis of Logic*. North-Holland. [2^a ed. Dover, 2006.]

Lambek, J., & Scott, P. J. (1986). *Introduction to Higher-Order Categorical Logic*. Cambridge University Press.

Lawvere, F. W. (1969). Adjointness in foundations. *Dialectica*, 23(3–4), 281–296.

Mac Lane, S., & Moerdijk, I. (1992). *Sheaves in Geometry and Logic*. Springer.

XIX. Psicologia del ragionamento e bias cognitivi

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press.

Johnson-Laird, P. N. (1983). *Mental Models*. Cambridge University Press.

Politzer, G. (2004). Reasoning, judgement and pragmatics. In I. A. Noveck & D. Sperber (Eds.), *Experimental Pragmatics* (pp. 94–115). Palgrave.

XX. Intersoggettività, embodiment e metafora concettuale

Gallese, V., & Lakoff, G. (2005). The brain's concepts. *Cognitive Neurodynamics*, 2(3–4), 455–479.

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. University of Chicago Press. [Trad. it.: *Metafora e vita quotidiana*. Bompiani, 1998.]

Lakoff, G., & Johnson, M. (1999). *Philosophy in the Flesh*. Basic Books.

Stern, D. N. (1985). *The Interpersonal World of the Infant*. Basic Books. [Trad. it.: *Il mondo interpersonale del bambino*. Bollati Boringhieri, 2005.]

Tomasello, M. (2009). *Why We Cooperate*. MIT Press.

Trevarthen, C. (1979). Communication and cooperation in early infancy. In M. Bullock (Ed.), *Before Speech* (pp. 321–347). Cambridge University Press.

XXI. Modelli formali dell'apprendimento e sistemi artificiali

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8), 1798–1828.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.

Negri, A. (2022). Logical and computational aspects of negation in large language models. *Preprint arXiv*, 2209.XXXXX. [Nota: risultati sperimentali sulla negazione nei LLM sono in rapida evoluzione; si rimanda alla letteratura più recente per aggiornamenti.]